INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

11TH MEETING

La Jolla, California (USA)
11-15 May 2020[1]

DOCUMENT SAC-11-07

YELLOWFIN TUNA IN THE EASTERN PACIFIC OCEAN, 2019: BENCHMARK ASSESSMENT

Carolina Minte-Vera, Mark N. Maunder, Haikun Xu, Juan L. Valero, Cleridy E. Lennert-Cody, and Alexandre Aires-da-Silva

CONTENTS

The staff's work, including the new assessments of tropical tunas, has been severely disrupted and delayed by the coronavirus pandemic, and many documents for the meeting of the SAC are not yet finalized. However, it is important that the members of the SAC and observers be informed as soon as possible of the direction and extent of the work, and of the very substantial progress that has been made, so some of the most essential documents are being published in draft form, and may be modified after discussion at the virtual sessions of the Committee.

EXECUTIVE SUMMARY

1. This year's benchmark assessment of yellowfin tuna in the eastern Pacific Ocean (EPO) is the basis for a risk analysis used to provide management advice (SAC-11-08). The risk analysis encompasses alternative hypotheses on the states of nature. The hypotheses were developed in a hierarchical framework that addressed uncertainties and issues with previous assessments. Previous assessments used a single model called a base case model, instead this assessment uses a set of reference models to represent the hypotheses.
2. The changes to the new assessment include:

[1] Postponed until a later date to be determined

  - improved index of abundance based on the dolphin associated purse -seine fishery;
  - new fishery definitions;
  - different assumptions about how the index relates to abundance;
  - different selectivities for some fisheries,
  - growth is estimated in some models.
- While previous assessments assumed a steepness of 1.0 to provide management advice, this assessment uses a total of 48 reference models, representing 12 different model configurations, each with four different values of steepness (0.7, 0.8, 0.9, 1.0).
- According to this assessment's 48-reference models, the point estimate for the spawning biomass of yellowfin at the beginning of 2020 ranged from 49% to 219% of the dynamic spawning biomass at MSY. The probability that the spawning biomass at the beginning of 2020 is lower than the dynamic MSY level is 50% or less for thirteen of the 48 models. The point estimate of the fishing mortality in 2017-2019 ranged from 40% to 168% of the fishing mortality at MSY. The probability that the fishing mortality of yellowfin in 2017-2019 is higher than the MSY level is 50% or more for fourteen of the 48 models.
- The point estimate for the spawning biomass at the beginning of 2020 ranged from 145% to 345% of the limit reference point. The probability that the spawning biomass at the beginning of 2020 is below the limit reference point ranges from 0 to 2%. The point estimate of the fishing mortality in 2017-2019 ranged from 22% to 65% of the limit reference point. The probability that the fishing mortality in 2017-2019 is higher than the limit reference point was estimated to be zero for all models.
- Every reference model with fixed lower steepness values results in more pessimistic estimates of stock status: lower spawning biomass relative to the target biomass reference point and higher fishing mortality relative to the target fishing mortality reference point. However, regardless of what value is assumed for steepness, some models estimate the stock to be overfished ($S<S_{MSY\_d}$) and overfishing is occurring ($F>F_{MSY}$) (models that assume either fixed growth, a linear relationship between the index of abundance and population abundance, no changes in selectivity through time, and asymptotic selectivity for the purse-seine fishery that catches the largest fish sizes). Conversely, other models estimate the stock is not overfished ($S>S_{MSY\_d}$) and overfishing is not occurring ($F<F_{MSY}$) (models that either assume dome-shape selectivity for the purse-seine fishery that catches the largest fish sizes). Stock status from the remaining models depend on the assumed steepness value.
- One of the key uncertainties in this assessment is about spatial structure. This is an avenue for future research to further improve the assessment models for EPO yellowfin tuna.
- The results from the reference models are combined in a risk analysis to provide management advice (SAC-11-08).

1. INTRODUCTION

This report presents the results of a benchmark stock assessment[2] of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean (EPO), conducted using an integrated statistical age-structured stock assessment model (Stock Synthesis version 3.30.15). It is the first assessment for this species undertaken by the Commission's scientific staff under the Work plan to improve stock assessments of tropical tunas and, although it uses the same modeling software platform as previous assessments, the methodology is quite different. The assessment now forms the foundation of a risk analysis, which seeks to explicitly take multiple sources of uncertainty into account in defining stock status and formulating management advice.

1.1.1. Background

Some problems and sources of uncertainty had arisen in the staff's assessment of yellowfin in recent years, leading to the staff not considering it reliable enough for management advice and ultimately its inclusion in the work plan to improve the stock assessments for tropical tuna in 2019. The main problem was that the yellowfin assessment results became overly sensitive to the inclusion of new data, in particular recent observations for the indices of relative abundance from the longline fishery (SAC-10 INF-F). As part of the work plan, a longline workshop and collaborative work with the main longline CPCs were conducted in 2019 to better understand and improve the longline data used in the bigeye and yellowfin assessments. As a result, the over sensitivity to the inclusion of the new data was found to be partially due to both the contraction of spatial extent and reduction in fishing effort of the Japanese fishery (whose data was used to estimate the longline index of relative abundance), resulting in a reduction in accuracy and precision of the index for recent years.

Other issues were identified which also related to the longline information such as a change in length composition data towards larger fish (SAC-10 INF-F) while the longline index showed a decline in recent years. The collaborative work between the staff and colleagues from longline nations suggested that this change may be due to changes in the fishery (*e.g.* gear or operation). However, the recent increase in the mean size of yellowfin is also seen in the catches of the dolphin-associated purse-seine fisheries and some unassociated purse-seine fisheries. This gives plausibility to the hypothesis that changes in some processes (*e.g.* selectivity changes) or model misspecification (*e.g.* growth) may be related to this increase. In some models of this benchmark assessment, growth and blocks in selectivity are estimated.

Another influential issue was inconsistencies between the longline and dolphin-associated purse seine indices of abundance**,** and the stock assessment model did not adequately fit them both. A new spatio-temporal modeling framework was developed and applied to the CPUE data to create new indices, but the inconsistencies remained unsolved.

One major source of uncertainty, and also a potentially explanation for the inconsistencies among indices is the possibility of spatial and stock structure of EPO yellowfin tuna that is not captured in the model. The staff's work for the 2nd external review of the yellowfin stock assessment included exploration of separate models for "southern" and "northern" hypothesized stocks. However, the review panel concluded that the "*evidence supporting a two-stock hypothesis was thought to be suggestive, rather than conclusive"*. In addition, that "*there was further evidence suggesting that YFT in the EPO was somewhere between a single, well-mixed stock and the two independent stocks"*. The review panel suggested various research avenues for the staff to explore to better account for stock structure in the yellowfin assessment (including one-stock and two stock hypotheses). While this benchmark assessment was conducted as if there

[2] "Benchmark" stock assessment means that model assumptions or methodologies previously adopted in this assessment were reviewed and improved if necessary, as opposed to an "update assessment", in which only the data used in the assessment have been updated.

were a single stock of yellowfin in the EPO, future research will continue to focus on alternative spatial structure hypotheses. To minimize potential biases resulting from ignoring spatial heterogeneity, this benchmark assessment is mainly fit to datasets representative of the area where the core of the catch is taken (north of 5°N). The catch taken south of 5°N are fully accounted for in the assessment, but the model does not fit to indices of abundance and length composition data from fisheries in the southern area (south of 5°N). This limits the influence of data that may be representative of another population unit.

There is also uncertainty about the nature of the stock-recruitment relationship (defined by steepness (*h*), which specifies how quickly recruitment decreases when the spawning is reduced). Previous assessments of EPO yellowfin tuna have consistently presented analyses of the sensitivity of the results to different assumptions about the *h* parameter, but only to show the impact of these assumptions on estimated management quantities; results of these analyses were not explicitly incorporated into the management advice. In this assessment, the uncertainty in *h* is explicitly included in the models.

Neither of the external reviews of the bigeye and yellowfin tuna assessments identified a replacement for the previous base-case models, but both review panels suggested a variety of alternatives for the staff to consider, particularly incorporating model uncertainty to derive information for management advice.

1.2. The new approach

This 2020 benchmark assessment of yellowfin tuna in the EPO, and the companion assessment of bigeye tuna (SAC-11-06), represent a new approach to stock assessments by the staff. Previously, a '*best assessment*' approach was used for the evaluation of stock status using a single 'base-case' model. The new assessments under the new approach are based on '*risk analysis*' methodologies, which use several *reference models* to represent various plausible *states of nature* (assumptions) about the biology of the fish, the productivity of the stocks, and/or the operation of the fisheries, and takes into account the different results, thus effectively incorporating uncertainty into the formulation of management advice[3]. This change, which represents a paradigm shift at IATTC, both for the staff's work and for the Commission's decision-making regarding the conservation of tropical tunas, also allows the staff to evaluate explicitly the probability statements specified in the IATTC harvest control rule for tropical tunas established in Resolution C-16-02.

This new approach to formulating advice for the management of tropical tuna fisheries includes the following four components:

1. Two **stock assessment reports**, for yellowfin (this document) and bigeye (SAC-11-06), describing data, model assumptions and presenting the results from all reference models for each species (model fits, diagnostics, stock status);
2. A **risk analysis** (SAC-11-08), assessing the consequences of using each model and the models combined as a basis for managing the fishery for tropical tunas by quantifying the probability of meeting the target and limit reference points specified in the IATTC harvest control rule;
3. **Stock status indicators** (SAC-11-05) for all three tropical tuna species (yellowfin, bigeye, skipjack); and
4. The **staff's recommendations** (SAC-11-15) for the conservation of tropical tunas, based on the above.

[3] See SAC-11 INF-F (Maunder *et al*. 2020a) for a description of the technical details of the risk analysis, using bigeye as a case study.

2. DATA

2.1. Fisheries and 'surveys'

Fishery-independent surveys are the gold standard to gather data to assess an exploited population. Due to their nature, most tuna fisheries worldwide have no surveys. The EPO in no exception, all data available to assess the stocks are obtained from the fishery. By a process of statistical standardization of fisheries-dependent data, however, an index of abundance and its associated length composition were estimated. Within the stock assessment models, the index and its length composition are treated as 'survey' data, thus modelled as having no catch and having a separate selectivity from the fisheries.

The fisheries are defined using several criteria, one of them is the geographical area of operation. This is consistent with the 'areas-as-fleets' approach, and it is a way to take spatial information into account without explicitly doing a spatial model.

The fisheries and 'surveys' defined for this assessment are illustrated in Figure 1, summarized in Table 1, and described in detail below.

2.1.1. Fisheries

Thirty eight fisheries are defined for the stock assessment of yellowfin tuna in the eastern Pacific Ocean (EPO), classified by gear (purse-seine, longline or pole-and-line), purse-seine set type (floating object, unassociated or dolphin), unit of catch (number or weight), quarter within a year and geographical area of operation (Figure 1, Table 1).

All the fisheries in this assessment, except the discard fisheries and the bait-boat fishery, use new spatial definitions based on spatial patterns in the length-frequency data identified with regression tree analyses (Lennert-Cody *et al*. 2010). The length-frequency data used in the regression tree analysis for purse-seine fisheries were obtained from the IATTC's port-sampling program (see Section 2.4. "Size-composition data") for 2000 to 2018. The length-frequency data for longline fisheries were the aggregated data (aggregated in space and time) from the Japanese fleet for 2000 to 2009 provided annually to IATTC by the National Institute of Far Seas Fisheries in Japan. The purse-seine data had a spatial resolution of 5° latitude by 5° longitude, and the longline data had a spatial resolution of 5° latitude by 10° longitude. The predictors for the tree analyses were quarter; cyclic quarter (*i.e.* cyclic combinations of quarters, such as quarters 1 and 4 *versus* quarters 2 and 3, quarters 1, 2, and 4 *versus* quarter 3, etc.); 5° latitude; 5° (10°) longitude. The response variable, which was multivariate, was the proportion of individuals in each length interval. The length intervals were 10-cm wide, with plus groups for the smallest fish and largest fish (≤ 39 cm; 40 – 49 cm; 50 – 59 cm; (…); 150 – 159 cm; ≥ 160 cm). For both the floating-object and longline fisheries, cyclic quarter (specifically, quarters 1 and 4 *versus* quarters 2 and 3) was found to be an important predictor for explaining variability in the length frequencies, and thus, seasonal fisheries were defined. Two longline fisheries, one in number and one in weight, are defined for each area and cyclic quarter combination, since longline catches are reported in number by some fleets and in weight by others.

The bait-boat fishery represents a small portion of the catches and has been declining over time, so it was treated as one homogeneous fishery for the whole EPO. The fisheries used to model discards have the same structure as in the previous assessment (SAC-10-07).

2.1.2. 'Surveys'

In Stock Synthesis, a 'survey' is modeled as a fleet that has data, such as indices of abundance and age/length compositions, but no catch. One 'survey' is used in this assessment was based on data from the purse-seine dolphin associated fishery (see section 2.3). Additional 'surveys' based on data from

longline fisheries were also defined and used in research in preparation for the benchmark assessment presented during the external review. The longline surveys are not fit in this assessment because their datasets are not representative of the core area where the catches are taken (north of 5°N).

2.2. Catch

The following types of catch data are defined for this assessment:

- **Retained**: catch retained aboard the vessel;
- **Discarded**: catches not retained aboard the vessel;
- **Total**: retained catch + discards;
- **Unloadings**: retained catch unloaded from the vessel.

2.2.1. Purse seine

The information used to estimate the total catch by species comes from four main sources (in order of importance): canneries, on-board observers, vessel logbooks, and in-port sampling by IATTC staff. If landing information from canneries is unavailable, catch information in the observer or vessel logbook databases, in that order, is used instead. The observer and logbook databases also contain other information about the catches, such as location and date caught, set type (on dolphin-associated tunas (DEL), on floating objects (OBJ), and on unassociated tunas (NOA) and vessel carrying capacity (<364 t (Classes 1-5) and ≥364 t (Class 6)); 'year' is the only ancillary information available in the unloading database. Additionally, since 2000, the port-sampling program for collecting length-composition data has also provided information on species composition (see section 2.4).

For this assessment, EPO total catches by species were estimated by catch strata, and then aggregated across catch strata to obtain quarterly estimates for each fishery. The catch strata are defined by area (those used in the stock assessment), month, set type, and vessel fish-carrying capacity. The method used to estimate the species composition of the catch depends on the sources of information available. Estimates prior to 2000 are based on the recorded species totals in the unloading or observer or logbook data, as applicable. To correct for underestimated bigeye catches, an adjustment factor that adjusts the catches of all three species, based on the port-sampling data from 2000-2004, is applied. The adjusted species totals are prorated to catch strata using the ancillary information in the observer and logbook databases. Since 2000, the port-sampling data have been used to determine the species composition of the total catch. The total catch of all three species combined (from unloading, observer and logbook data) is prorated to catch strata, using the information in the observer and logbook databases. The port-sampling data on the species and size composition of the catch are then used to estimate the catch of each species by catch stratum. Detailed explanations of the sampling and estimators can be found in the appendix of Suter (2010) and in WSBET-02-06.

2.2.2. Longline

The IATTC staff does not directly collect data on longline catches; these are reported annually to the IATTC by the Members and Cooperating non-Members following the resolution C-03-05 on Data Provision. Catches are reported by species, but the availability and format of the data varies among fleets. The principal fleets report catch and effort aggregated by 5° cell-month. IATTC databases include data on the spatial and temporal distributions of longline catches of yellowfin tuna in the EPO by the fleets of distant-water CPCs (China, Chinese Taipei, French Polynesia, Japan, Korea, Vanuatu) and coastal CPCs (principally Mexico and the United States).

For this assessment, these data are aggregated according to the new fishery definitions based on area of operation, and cyclic quarter combination (Figure 1, Table 1). For each area and cyclic quarter combination (quarters 4, 1 *versus* quarters 2,3), two fisheries are defined to enter the catches in their original reporting

units (number or weight), and the conversion from numbers to weight is done internally in the assessment model.

Updated and new catch data for the longline fisheries (Fisheries 29-40), available to the IATTC staff as of April 10, 2020, were incorporated into the current assessment. New or updated catch data were available for Vanuatu (2018), Chinese Taipei (2016- 2018), French Polynesia (2018), China (2018), Japan (2016-2018), Korea (2018) and the United States (2014-2018). For 2019, and for other years when catch may not be available, catches were set equal, by CPC, to the last year for which catch data were available. For fleets that reported catch aggregated by year and space, the data was disaggregated using the proportion of catches by quarter and area for the closest year for which data on that resolution were available. The catches of a coastal CPC that reported aggregated catches were added to the area which contained that CPC's Exclusive Economic Zone (EEZ). The algorithm to calculate the catches is described in WSBET-02-03,

2.2.3. Discards

Two types of discards are considered, those resulting from inefficiencies in the fishing process and those related to the sorting of catches. Examples of inefficiency are catch from a set exceeding the remaining storage capacity of the fishing vessel or dumping unwanted bycatch species, while catch sorting is assumed to occur when fishers discard tuna that are under a certain size. Discards by the longline fisheries cannot be estimated with the minimal data available (given small % of observer coverage), so it is assumed that the retained catch represents the total catch (Table 1). Both types of discard by purse-seine are estimated.

The amount of yellowfin discarded, regardless of the reason, is estimated with information collected by IATTC or national observers using the methods in Maunder and Watters (2003). No observer data are available to estimate discards prior to 1993, and it is assumed that there were no discards before that time. Also, there are periods for which observer data are not sufficient to estimate the discards, in which case it is assumed that the discard rate (discards/retained catches) is equal to the discard rate for the same quarter in the previous year or, if quarterly data are not available, a proximate year. Removals by Fisheries 1-10 (purse-seine on floating objects) are retained catch plus some discards resulting from inefficiencies in the fishing process. The removals by Fisheries 11-14 (purse-seine unassociated) are retained catch, plus some discards resulting from inefficiencies in the fishing process and from sorting the catch. Discards that result from the process of sorting the catches in the floating-object fisheries are treated separately (Fisheries 25 to 28), following the rationale of Maunder and Watters (2001). These discards are assumed to be composed only of fish that are 1-3 quarters old. Sorting is infrequent in the other purse-seine fisheries.

2.2.4. Catch and discards trends

Yellowfin tuna has been fished in the EPO since the early 1900's (Estes 1983). Prior to the 1950's, the fishery occurred mostly within 250 miles of the coast or around islands and sea mounts, and was done mainly by bait boats, followed by purse seiners (Peterson and Bayliff 1985). In the 1950's the longline fisheries started expanding in the EPO from the WCPO, getting to the coastal areas around the mid 1960's, mainly targeting bigeye tuna, but catching yellowfin tuna as a secondary species (Shimada and Schaefer 1956; Matsumoto and Bayliff 2008). The catches of baitboats have been small and declining over time.

The purse-seine fisheries, mainly associated with dolphins, became the main fishing method in the 1960's, and has continue since (Figure 2). The main dolphin-associated fisheries are close to Central America and southern Mexico (F18-DEL_C, F19-DEL_*P*; Figure 1). The purse-seine fishery associated with floating objects has been important since the 1970's in areas north of the equator (OBJ-N and OBJ-Nc, Figures 1 and 2) and close to the coast of South America, between 10°S and the equator (OBJ-Cc). The fisheries on floating objects had a widespread expansion in the EPO after 1992. In the last 15 years, the number of sets on

floating objects has been steadily increasing towards its highest current levels. (SAC-11-05). Catches from areas C and S increased after 1992. The main unassociated purse-seine fisheries have been in the north (NOA-N) and, after 2010, in the central EPO (NOA-C). The discards due to sorting in the floating-objects fisheries shows a reduction beginning around 2001, following resolutions adopted by the IATTC during 2001-2007, which prohibited discarding of small tunas.

Longline catches represent a small proportion of the total catches of yellowfin tuna in the EPO (Figure 2). The main longline fishing areas have always been in the western EPO, where a decline occurred in the late 2000's, but a new increasing trend is apparent since 2010. The longline catches in the eastern EPO (LL-E) area are characterized by a marked seasonality and have declined in recent years.

2.3. Indices of abundance

While both purse-seine and longline indices of abundance are available for yellowfin in the EPO, this assessment includes only the purse-seine index. There was an inconsistency between the index based on Japanese longline CPUE and the indices based on dolphin associated purse-seine CPUE. Extensive work was done in collaboration with the longline CPCs to better understand the data, incorporate new data, and conduct new analyses. A longline workshop was conducted and scientists from Japan and Korea collaborated with the staff to further address the issues. A new spatio-temporal modeling framework was developed and applied to the CPUE data to create new indices (Xu *et al*. 2019) but the inconsistencies remained unresolved. Therefore, it was concluded that stock assessment model cannot adequately fit both types of indices simultaneously. To deal with this issue, the purse-seine index was selected for this benchmark assessment. This is because the longline catches represent a small proportion of the yellowfin catches, and the distribution of the Japanese fleet has been contracting towards the western EPO, away from the "core" catch areas for yellowfin (see Figure A1 in SAC-11-06). It is noted that, given these spatial changes, one potential explanation for the inconsistencies among indices in the model is the possibility of spatial structure in the EPO yellowfin population.

2.3.1. Data selection

The data used to construct the index are the set-by-set catch and effort observations from purse-seine vessels. The data were collected by onboard observers of the Agreement on the International Dolphin Conservation Program (AIDCP) observer program or obtained from vessel logbooks. Observer coverage of the purse-seine fishery has been largely limited to large vessels (fish-carrying capacity> 363 t) and has increased over time to at or nearly 100% since 1992 (Joseph, 1994; Scott *et al*., 2016). Logbook data were used for trips of large vessels for which no observer data were available.

The data selection was done to retain only those vessels with similar fishing strategies each year using two selection criteria. First, a vessel was only selected in years when at least 75% of its annual sets were made on tunas associated with dolphins. Second, among the vessels that met the first criterion, those represented by at least 10 years of data (with more than 75% of sets on dolphins) and over at least 18 years of data coverage (the difference between the first and last year of data available for the vessel) were considered dolphin-associated vessels and their data retained for the index standardization. This data selection procedure is necessary because it is not possible to separate searching effort by set type. Only the main fishing grounds were selected for the standardization by including the 1° sampling cells that were represented by at least 30 years of data between 1985-2019, and were located north of 5°N. A total 52 vessels were retained after the data selection to be included in the standardization (Figure A1). The effort unit used was the number of days fishing.

2.3.2. Standardization procedure

The standardization of the catch and effort data for yellowfin tuna from the dolphin-associated purse-

seine fishery was conducted using the R library *VAST* (version 3.0.0) (Thorson and Barnett 2017, Xu *et al* 2019, Maunder *et al* 2020b). VAST fits a delta-generalized linear mixed effects spatiotemporal model to data. It models separately the encounter probability and positive catch rate, which are assumed to have a logit and log link, respectively, and combines the results to produce the final estimates. There are several advantages of using mixed-effects spatiotemporal models over the fixed-effects generalized linear models conventionally used in CPUE standardization. First, the estimation of spatiotemporal correlations allows for the prediction of catch rates in unfished locations based on the information from neighboring areas/times. Second, the uncertainty estimates take into consideration the spatial coverage and sample size. Third, the final estimates are naturally weighted the by the area related to each knot in the spatial domain, rather than by the sample size. Both the encounter probability and the catch-rates are modeled with linear predictors that include an intercept term (year-quarter effect), vessel effects on catchability and a spatial effect (Xu *et al*. 2019). The spatial effect is represented by a mesh of 200 knots. The model converged (gradient=0.0004) with a positive definite Hessian.

The index shows four noticeable periods (Figures 3 and 4):

- High abundance from 1984 to 2000, and a high abundance in 1996
- A marked increase in abundance in 2001 that persisted to the middle of 2003
- A period of decreased abundance until the beginning of 2015
- The recent period of further lower abundance

The lower spatial coverage and sample size in the early years (Figure 4) resulted in a higher CV for those years.

2.4. Size-composition data

2.4.1. Fisheries

2.4.1.A Purse-seine

The length-frequency data for the purse-seine fisheries are obtained through the port-sampling program of the IATTC. The sampling is done by IATTC personnel at ports of landing in Ecuador, Mexico, Panama, and Venezuela. The ancillary information available in the port-sampling database is determined by the port-sampling protocol (Tomlinson, 2002; Suter, 2010), where samples are collected according to sampling strata. Sampling strata are defined by the fish-carrying capacity of the vessel, the date (12 months), the type of purse-seine set (sets on dolphin-associated tunas; sets on floating-object-associated tunas; sets on unassociated schools of tunas) and area of fishing (13 areas; see Figure 1 in WSBET-02-06). Wells are the primary sampling unit within a stratum, with unequal numbers of wells sampled per stratum. Fish within a well are the secondary sampling unit. Sampling at both stages is largely opportunistic with the exception that wells are sampled only if all the catch within the well came from the same stratum. Analysis of data from recent years has shown that this restriction can result in a predominance of large-catch sets that are sampled (Lennert-Cody and Tomlinson 2010). More than one well may be sampled per vessel if the catch of other wells comes from different strata. However, typically only one or two wells per trip are sampled. About 50%-60% of trips of large purse-seine vessels and 10%-20% of trips of small purse-seine vessels have typically been sampled by the port-sampling program per year, leading to a total of over 800 wells sampled in most years (IATTC 2010a; Vogel, 2014). The percentage coverage of the catch is lower (SAC-02-10). The sampling areas were designed for yellowfin prior to the development of the fishery on fish-aggregating devices (FADs) and have remained the same. Since 2000, the 5° area, in addition to the sampling area, was recorded for almost all samples (Lennert-Cody *et al*. 2012). Prior to 2000, the 5° area has been recovered for many samples. Ideally, 50 fish of each species in the catch of the sampled well were measured. Since 2000, samplers alternate between counting fish by species and measuring fish for length (fish included in the counts are independent of fish measured).

As with the species catch, the size composition of the catch is estimated by catch strata and then aggregated across catch strata to obtain quarterly estimates for each fishery. The estimates are in numbers of fish by 1-cm length interval, which are then converted to proportion of fish at length for the assessment. The estimated numbers at length are obtained by multiplying the well-level estimates of the proportion at length, combined across sampled wells, by the estimated total catch in numbers for the species in the stratum. Since 2000, the well estimates of proportions at length make use of both the species counts and the length measurement data. Details of the estimators can be found in WS-BET-06-02.

The fisheries that catch the smallest sized fish are the bait boat (pole and line) and the floating object purse-seine fisheries (Figures 5A and 5B). The fisheries that catch the largest sized fish are the dolphin associated purse-seine fisheries and the longline fisheries. Trends over time are seen in the average size fish caught by the purse-seine fisheries (Figure 5B).

2.4.1.B Longline

The length-composition data for longline fisheries are used in preliminary runs to obtain the best fit for the asymptotic selectivity of the longline fisheries. The models composing the reference table are not fit to these data because of the potential spatial structure of the yellowfin tuna population in the EPO and the assessment is focus on the portion of the population exploited by the purse-seine fishery.

The length-composition data for longline fisheries in this assessment are based on 1) new monthly 1° x 1° length-frequency data for the Japanese commercial fleet; 2) new monthly 1° latitude x 1° longitude catch and effort data for individual Japanese commercial vessels; and 3) 5° latitude x 5°longitude quarterly longline catch data reported by CPCs. The length-composition data should be representative of all longline catches, so the monthly 1° cell length-frequency observations are raised to the fishery catch in a quarterly 5° cell in the same strata, as follows:

1. Raise monthly Japanese 1° x 1° length-frequency data (with 1 or 2 cm resolution) to total Japanese catch in the same strata;
2. Aggregate the raised data from step 1 to quarterly 5° x 5° catch;
3. Raise the aggregated length-frequencies from step 2 to the total catch of all CPCs in the same strata;
4. Aggregate the raised length-frequencies from step 3 in the longline fisheries 29 to 35 of Table 3.1

The length compositions, from 20 cm to 198+ cm, are aggregated at 2-cm intervals, and their input sample sizes are computed as the total number of fish sampled divided by 100. The input sample sizes for every longline fishery have decreased continuously since the mid-1990s and reached very low values in most recent quarters due to the decline in the catches in the Japanese fleet and the switch in sampling strategy from crew samplers to observers onboard, and thus fewer fish being sampled.

2.4.2. Survey

The length frequencies of yellowfin tuna associated with the index of abundance were also obtained from the standardization of the data from the dolphin-associated purse-seine fishery using VAST, with the inclusion of a multivariate response variable (Thorson and Haltuch 2018, Maunder *et al*. 2020b). The data used were the length frequencies collected by the port-sampling program. The length frequencies raised to the well catch, were aggregated by quarter, 5° cell and set type aggregated. The aggregated data was raised to the catch in a stratum using data from the observer and logbook databases. Strata were defined as quarter-5°cell combinations. The vessel and spatial cell selection criterion was the same as the CPUE. The multivariate response variable was length specific catch rate (in ton day-^{1} fished). The length frequency classes were defined by 10 cm intervals, from 20 to 190 cm.

The models treats the encounter probability and positive catch rate separately, with logit and log links, respectively. The linear predictors are spatial and the temporal (year-quarter) components. The spatial component is represented by 30 spatial knots (that aggregate the 5°cells to improve computational efficiency). The sum of the indices by length class were similar to the overall index (section 2.3), indicating that the standardized length frequencies are a good representation of the length classes vulnerable to the index of abundance. The model converged (gradient=0.000006) with a positive definite Hessian.

The classes with largest frequencies ranged from 40 to 160 cm (Figure 6) with most lengths within 70 to 120 cm, except in two periods: 1) from 2002 to 2007, when an increase in the proportion of small-sized fish (<70 cm) was maintained for several quarters in a row; 2) from the end of 2014 onward, when the proportion of large-sized fish (>120cm) increased and stayed high (Figure 6).

2.5. Age conditional on length

Age and length data (Wild 1986) were used to provide information when growth is estimated in the stock assessment model. Wild's data consist of ages, based on counts of daily increments in otoliths, and lengths for 196 fish collected between 1977 and 1979. The sampling was conducted by collecting 15 fish in each 10-cm interval in the length range of 30 to 170 cm and are therefore included in the model as conditional age-at-length. For the largest size ranges, Wild was unable to complete the sample size of 15 fish, due to the scarcity of fish those sizes or unreadability of the otoliths (Wild 1986). The daily periodicity of the rings has been validated for fish from 25 to 146 cm (Wild and Foreman, Yamanaka 1990, Wild *et al* 1995) and in larvae up to 16 days after hatching (3-7 mm standard length) (Wexler *et al*. 2001). The maximum age obtained was 4 years old (Wild 1986). In the model, the data was included as age frequency (quarters) aggregated by length class and not disaggregated by sex, and it was assumed that the data were sampled from fishery F18-DEL_C, a dolphin associated purse-seine fishery that catches a wide range of fish sizes (Figure 5A), and at the beginning of the period modelled (1985). Figure 7 shows the age frequency conditional on length class and the fixed assumption for growth (see 3.1.1)

3. ASSUMPTIONS AND PARAMETERS

An integrated statistical age-structured stock assessment model was used to do the benchmark assessment of yellowfin tuna in the EPO for 2019 (*Stock Synthesis* Version V3.30.15;_2020_03_26, Methot and Wetzel 2013). Two subsequent unreleased versions (V3.30.15.03-opt and V3.30.15.04-safe, provided by Rick Methot, NOAA Fisheries) were used to estimate the variability of F_{cur}/F_{MSY}, F_{cur}/F_{LIMIT} and dynamic S0. Francis weights, recruitment deviation bias corrections and other auxiliary quantities and graphs were produced using the *R* library *r4ss* (version 1.38.0), the set of packages from *tidyverse* (1.3.0) and original code available from the IATTC repository IATTCassessment.

The first year of the model is 1984 and the last 2019. The start year differs from the previous stock assessments, which started in 1975, because data from the purse-seine fishery with spatial information necessary to standardize the index and length frequencies was limited before 1984. The time step of the model is a quarter, 30 age classes are defined, from 0 quarters to 29+ quarters (7.25 years). The population size structure was defined as 2-cm intervals from 2 to 200+ cm. The model is structured by sex, with sex-specific natural mortality. The size composition are defined using 2-cm bins classes from 20 to 198+ cm for the fisheries and 10-cm for the 'survey', from 20 to 190 cm. The models are fit to catches (with high precision), relative abundance indices, and size composition data. Models that estimated growth were also fit to age conditioned on length data. The observed total catches were assumed to be unbiased and relatively precise and were fitted assuming a lognormal error distribution with standard error (SE) of 0.05.

3.1. Biological and demographic information

3.1.1. Growth

The average length at age is assumed to follow a Schnute-Richard's curve (Richards, 1959; Schnute, 1981) reparameterized with L_1, L_2, a_1, a_2 as implemented in Stock Synthesis (Methot and Wetzel, 2013). The age a_1 is the first age that the fish start following a Schnute-Richards curve, and the corresponding mean size at this age is L_1. The growth curve is adjusted to go through the size L_2 when the fish are age a_2 (Methot, 2012):

$$L_a = L_1^b + (L_2^b - L_1^b)\left(\frac{1-\exp(-K(a-a_1))}{1-\exp(-K(a_2-a_1))}\right)^{1/b} \quad \text{(Equation 1)}$$

There is uncertainty in the growth of yellowfin tuna in the EPO. Wild (1986) estimated L_∞ at 188.2 cm using a Richards curve fit to otolith age-length data. This estimate, however, is an extrapolation well beyond the maximum age of 4 years in Wild (1996) study. A limited set of reliable, but restricted in its spatial and temporal distribution, tagging data is consistent with an L_2 of about 172 cm.

Maunder and Aires-da-Silva (2009) estimated growth within the stock assessment model and these parameter estimates have been used in previous assessment and are used in this analysis form models that assumed fixed growth: L_1= 18.3686, L_2 = 182.307, a_1 = 0 quarter, a_2 = 29 quarters, K = 0.19228628 quarter-1, b = -0.542255.

Misfit of the length composition data (mainly to the fishery with asymptotic selectivity) from preliminary runs with fixed growth indicates that the assumed growth function may not represent growth for the core of the exploited population. Given the uncertainty in growth, models were included in the reference set that estimate growth while fitting to the conditional age-at-length data.

The variability of size at age may also be important as this value will determine what sizes are plausible to exist in the population. In this assessment we assumed a 7.5% CV for the variability of size at age for ages. This value was set as a compromise between the former stock assessment assumption of 10%, which was considered too large during the external review and the assumption of 5% used in research models presented at that review, which was considered too small. Figure 7 shows the fixed growth curve and the variability assumption; 95% of five-year-old fish (20 quarters) have sizes ranging from 144 to 193 cm, 7.25-year-old fish (a_2) have sizes between 155 and 209 cm.

The weight at age w_a is obtained by replacing the average length at age L_a in the length-weight equation for yellowfin tuna in the EPO (Wild, 1986):

$$w_a = 1.387 \times 10^{-5} L_a^{3.086} \quad \text{(Equation 2)}$$

3.1.2. Natural mortality

For this assessment, as in previous assessments, it is assumed that, as yellowfin grow older, the natural mortality rate (M) changes (Maunder and Aires-da-Silva 2012). Males and females are treated separately in this assessment, and M differs between males and females (Figure 8). The largest natural mortality is at age zero, then declines almost linearly until they are 30 months old, then increases again for females These values were estimated by fitting to sex ratio-at-length data (Schaefer 1998), and comparing the values with those estimated for yellowfin in the western and central Pacific Ocean (Hampton 2000; Hampton and Fournier 2001). Maunder and Watters (2001) describe in detail how the age-specific natural mortality schedule for yellowfin in the EPO was determined.

The assumed level of natural mortality for age zero has no impact on the assessment results. Recruitment

occurs at age zero in the assessment model. Age zero is used for convenience, and the assumed natural mortality for ages not vulnerable to the fisheries is not intended to represent the actual natural mortality, and only arbitrarily scales the recruitment at age zero.

3.1.3. Reproductive biology and recruitment

Yellowfin tuna can spawn almost every day if the water temperatures are in the range of 24 to 30°C, resulting in spawning year-round in lower latitudes and in the summer in higher latitudes (Nishikawa *et al*., 1985; Schaefer, 1998; Itano, 2000). An "index" of total egg production (fecundity), rather than the spawning biomass, is used in the assessment. This is obtained from the reproductive biology study by Schaefer (1998) (Figure 9). The fecundity O_t^F at time t at age a is given by:

$$O_t^F = \sum_{a=0}^{29} \frac{p_a f_a d_a N_{a,s=1,t}}{1{,}000{,}000} \quad \text{(Equation 4)}$$

where p_a is the proportion of mature females at age a, f_a is the batch fecundity at age a, d_a is the fraction of females spawning per day at age a. To obtain the p_a, f_a, and d_a, from those quantities estimated at length by Schaefer (1998), the average length at age was used in the equations below.

The proportion of mature females at length p_l is :

$$p_L = e^{-(\exp(-0.059347(L-85.901241))} \quad \text{(Equation 5)}$$

where *L* is the fork length in centimeters.

The batch fecundity (number of migratory-nucleus or hydrated oocytes in the ovary) for a female with fork length (in millimeters) is

$$f_L = 0.0003747\, L^{\,3.180758} \quad \text{(Equation 6)}$$

The fraction of females spawning per day d_l at fork length L (in centimeters) is:

$$d_L = 0.742(1 - e^{-0.046(L-54.892)}) \quad \text{(Equation 7)}$$

Four recruitments are estimated in a year. Recruitment (age zero fish) is assumed follow a Beverton and Holt (1957) stock-recruitment relationship. The Beverton-Holt curve is parameterized so that the relationship between spawning biomass (fecundity in this assessment) and recruitment is determined by the average recruitment produced by an unexploited population (virgin recruitment) and steepness (*h*). Steepness is defined as the proportion of the virgin recruitment that a population produces when reduced to 20% of its virgin state. A steepness of 1 implies that the stock may produce recruitments equal to the virgin level, on average, at all levels of spawning biomass, while a steepness of 0.70 indicates that when a stock is at 20% of its virgin spawning biomass, only 70% of the virgin recruitment is produced, on average.

Steepness is a key parameter of a stock assessment, but it is problematic to estimate (Lee *et al*. 2012). For tunas, there is little evidence for any particular value. In previous assessments the base-case model had the assumption of *h*=1. This assessment incorporates the uncertainty in steepness by including four hypotheses in the reference models, *h*=0.7, 0.8, 0.9, or 1.0.

The recruitment is assumed to vary lognormally around the stock recruitment curve with a standard deviation of 1 on the logarithm of the recruitment deviations. The variability of the recruitments is constrained by a penalty added to objective function. The recruitments are corrected so that the expected values are unbiased. The bias correction is computed using the method of Methot and Taylor (2011).

3.1.4. Movement and stock structure

Yellowfin is widely distributed in the tropical and subtropical waters of the Pacific Ocean. Yellowfin are

found principally in the mixed layer at temperatures between 20°C and 30°C but may perform "bounce" dives below the thermocline for foraging during the day (Schaefer *et al.* 2007). Juveniles and small fish may aggregate around floating objects while older fish may be found associated with several species of dolphins. Although considered a highly migratory species, the tagging studies done in the EPO have indicated that yellowfin tuna move in restricted areas mostly within 1000 nautical miles of their tagging locations (Fink and Bayliff, 1970; Bayliff, 1979, 1984, Schaefer *et al*., 2011; Schaefer *et al*., 2014). However, the evidence from tagging data is not enough to support neither complete mixing nor spatial (Joseph *et al* 1964, Schaefer 2009). Genomic studies are promising in detecting stock structure, and in the Pacific Ocean some genomic evidence for heterogeneous structure exists (Grewe *et al* 2015, Pecoraro *et al* 2018). No such study has been done within the EPO. While yellowfin tuna in the EPO may be composed of spatially disaggregated units (Schaefer, 2009) the available data is insufficient to estimate movement rates or assist in the delimitation of these units. For this assessment, as in previous assessments, it is assumed that there is a single stock of yellowfin tuna in the EPO.

3.2. Fisheries dynamics

3.2.1. Initial conditions

The model is assumed to start from a non-virgin (fished) equilibrium state, with R_{init}, the initial recruitment as an offset of the virgin recruitment, and F_{init}, the initial fishing mortality, being estimated, with no penalty on initial equilibrium catches, thus allowing for unconstrained estimation of the initial conditions. The initial fishing mortality was assumed to correspond fleet F16-DEL_NE. This fleet was chosen because it catches a wide range of sizes, thus it could represent best the equilibrium fishing mortality at age for the stock. Additionally, 16 recruitment (quarter) deviations before the start of the model initial quarter are estimated.

3.2.2. Selectivity

The selectivity was modelled as a function of length and age, except for the "discard" fisheries, for which only selectivity at age was assumed (fixed at 1 for ages 1 to 3 quarters, and 0 otherwise). For all other fleets, the selectivity was fixed to be 1 for all ages, except for age zero, which had selectivity of zero.

Selectivity curves at size were assumed to be dome-shaped for most fleets and were modeled initially using double-normal functions. The preliminary fits to the double-normal function were unsatisfactory and indicated that more flexible selectivity functions should be used. The need for more complex and flexible shapes for selectivity may be because the selectivity encompasses not only the gear selection pattern but also the spatial-temporal availability of fish of different sizes. It is likely that seasonal patterns are present because of oceanographic conditions or movement, and those will be absorbed in the selectivity curves. Therefore, selectivity curves based on cubic splines were adopted (Table 2). The cubic splines' number and location of nodes was initially obtained by fitting splines to the ratio of fish in the catches and the numbers in the population ("empirical" selectivity), at length, obtained from preliminary models using double-normal selectivity functions. The fits were done using an external fit in *R* (library freeknotsplines). Initial fitting was performed with the suggested spline configurations and then fine tuning was done. A node at the beginning and another at the end of the size distribution for the fleet was always included to avoid extreme changes in selectivity at the tails of the distribution. The parameter for one of the splines nodes was fixed at an arbitrary value and the values for the other nodes were estimated relative to that fixed parameter.

The selectivity for the longline fleets and were assumed to be asymptotic and fixed to values estimated in preliminary runs. The selectivity for the purse-seine fisheries south of 5°N were set equal to those of the longline fisheries. The composition data for those fisheries were not fit in the reference models, only in the preliminary runs. The selectivity for the longline fisheries in weight were "mirrored", that is

constrained to have the same parameter estimates, to the equivalent fishery in numbers.

The basic assumption is that selectivity is time-invariant. There are two periods that seem to depart from this assumption (Figures 5B and 6). The first period was during, and for some quarters after, the peak in the index of abundance in about 2002 (Figure 3). The second period occurred in the recent years, since about 2015, when the size of yellowfin in this fishery was higher on average. Because the standardize size compositions associated with the index (Figure 6) are derived from dolphin-associated data, those changes in average size are also seen in the index. These uncertainties regarding selectivity were addressed in the reference set.

The dolphin-associated purse seine fishery that catches the largest yellowfin in the core area (F19-DEL_*P*) was chosen as the fleet with asymptotic selectivity for which its length composition was included in the objective function. Models with asymptotic selectivity had poor fits to this fisheries length composition data. To improve the fit, an alternative hypothesis that this fishery has dome-shaped selectivity was also considered in the reference set.

3.3. Data weighting

Likelihood functions encompasses not only the sampling (observation) variability, but also model misspecification and unmodelled process variability. Therefore, the CVs of the index of abundance are set equal to the CVs estimated from the standardization model plus a constant added such that the average CV for a range of years is 0.15 (the average CV for the whole time series about 0.18).

The size composition data were assumed to have multinomial distributions with the variance proportional to the sample size. The input sample size for the purse-seine fisheries was equal to the number of wells sampled. The number of fish sampled within a well cannot be used to represent the sample size because fish stored in the same well may come from the same school and thus are not independent samples and their sizes may be highly correlated (Pennington *et al*. 2002). For the preliminary model runs used to estimate the selectivity of the longline fisheries, the length frequency of the longline fishery was used and the sample size was set equal to the number of fish/100. The Francis method for weighting the size composition data was used (TA1.8 in Francis 2011). A preliminary run was conducted with weighting equal 1 and reweighting factors ("Francis weights") were computed based on how well the model fitted the size length composition data (Table 2). In addition to that, all length compositions with multimodal distributions were further downweighed by multiplying the Francis weights by 0.5, since it is likely that due to processes not model explicitly (*e.g.* movement). Similarly, to the index of abundance, the length composition likelihood will also absorb model misspecification and unmodelled process variability.

3.4. Model diagnostics

A suite of approaches was used as diagnostics to determine whether a model fits the data well and is correctly specified:

Index of abundance: The root-mean-square error (RMSE) of the residuals was compared to the input CV to evaluate how well the reference models fit the index of abundance and to evaluate the validity of the variability assumption. The residuals are inspected for trends or patterns that may indicate model misspecification.

Size composition data: predicted and empirical selectivity curves for every fishery that has composition data were compared. The empirical selectivity of a fishery is defined as the average observed catch at length from the fishery divided by the average predicted population number at length. The empirical selectivity was scaled to a maximum value of 1, unless noted otherwise. If the assessment model fits a fishery's composition well, the two selectivity curves should be similar. The residuals of the length composition data were inspected for trends over time and across length classes. The effective sample size

(McAllister and Ianelli, 1997) implied by the model fit was compared to the input sample size. The effective sample size is the size of the random sample needed, on average, to achieve a fit that is as good as the variance in the model's fit to the composition vector (Methot and Wetzel 2013). The better the fit, the larger the effective sample size.

Integrated model diagnostics: Age-structure production models (Maunder and Piner 2015), catch-curve analysis (Carvalho *et al* 2017), likelihood profile on the global scaling parameter (Lee *et al* 2014, Wang *et al* 2014) and retrospective analyses (Mohn 1999, Hurtado-Ferro *et al* 2015) were used to detect model misspecification, influence of different data sets, and other potential issues with the models (Appendix A2).

4. REFERENCE MODELS

This benchmark assessment is the basis for a risk analysis that addresses the uncertainties about several assumptions and explicitly includes uncertainty in the evaluation of stock status and formulation of management advice (SAC-11-08).

The first step to apply the risk analysis framework (SAC-11 INF-F, Maunder *et al*. 2020a) is to list the unresolved issues and uncertainties that need to be accounted for in the management advice. Then several hypotheses are formulated that represent different states of nature that may resolve these issues or represent the uncertainties, and these are arranged in a hierarchy . The most encompassing hypotheses (overarching hypotheses) are at the top of the hierarchy, then other levels unfold nested under the upper levels. The main issues and uncertainties when assessing the stock status of yellowfin tuna include: a) spatial structure; b) inconsistencies between the indices of abundance based on CPUE from the dolphin-associated purse-seine fishery and that based on CPUE from the longline fishery; c) inability of the model to fit the high values in the indices of abundance; d) and misfit to the composition data for the fishery that is assumed to have asymptotic selectivity.

The overarching set of hypotheses (Level 1, Figure 10a) addresses the issue of spatial structure. Although there is some evidence of northern and southern stocks, the divisions are not clear and mixing between the two potential stocks may be episodic or the magnitude may be variable from year. The overarching hypotheses for the spatial structure of yellowfin tuna in the EPO formulated were "High mixing", "Episodic/high variability mixing", and "Negligible mixing". The "High mixing" overarching hypothesis is represented by single-stock models similar to previous assessments. The "Episodic/high variability mixing" overarching hypothesis is represented by single-stock models that are driven by the north or the southern stock data. This means that the model is fit to data for the north (south) and only uses the catch for the south (north) while fixing the selectivity for these fisheries. The "Negligible mixing" hypothesis is represented by two independent assessments, one for the north and one for the south. Many of these models were developed for the yellowfin tuna review and this informed the decision to eliminate all hypotheses except the "High mixing" hypothesis from the risk analysis to make it practical to implement. This assessment thus focuses on the hypotheses nested within the overarching "High mixing" hypothesis (Figure 10b).

Under the High mixing hypothesis in Figure 10b are hypotheses that address the misfit to the index of abundance and the changes in selectivity (Level 2A), and the misfit to the length composition data for the fishery with asymptotic selectivity (Level 2B). Models representing different steepness scenarios are added as a third level in the hierarchy (Level 3). There are four hypotheses at Level 2A to address issues with the index misfit and changes in selectivity:

Index is proportional to abundance **(BASE):** This model is most similar to previous models used to assess yellowfin tuna in the EPO and is the basis for all the other models. This model assumes that the index is proportional to abundance for the whole time period.

Density dependent catchability (**DDQ**): This model allows the estimation of a coefficient *c* that determines how catchability is influenced by abundance. It is hypothesized that during periods of high abundance the purse seine fleet that fishes on yellowfin associated with dolphins can more efficiently catch yellowfin tuna and this will allow the model to better fit the high index observations (hyper-depletion, $c > 0$, Methot *et al* 2020).

Time block in the middle (**TBM**): This model allows the estimation of a block in catchability and selectivity of the index during the period where there is peak abundance. Around this period and shortly after, the fishery associated with dolphins catches smaller yellowfin on average. It is hypothesized that during the period of high abundance the purse seine fleet that fishes on yellowfin associated with dolphins can more efficiently catch yellowfin tuna and this will allow the model to better fit the high index observations, but assumes that at other times this is not the case. It assumes that if catchability changes, selectivity is also likely to change.

Time block at the end (**TBE)**: This model allows the estimation of catchability and selectivity for the survey and some fisheries during the later period where the size of fish caught by the purse seine fishery associated with dolphins is higher on average. It is hypothesized that if selectivity for the index changes, catchability is also likely to change.

There are three hypotheses at Level 2B that address the misfit to the length composition data:

Fixed growth (BASE): This is the same model as above, it represents the null hypothesis that growth is well described using the fixed parameters.

Estimate growth (**GRO**): The fixed value used for asymptotic length is higher than the limited tagging data, but is somewhat consistent with the otolith data, although old fish cannot be aged using the daily increment method. The otolith data comes from before the model starts and the tagging data is limited in its spatial and temporal distribution. Therefore, estimation of growth within the stock assessment model may be appropriate and may allow a better fit to the length composition data for the fishery with asymptotic selectivity. The age conditioned on length data is also included in the model to inform the estimates of growth.

Dome-shape selectivity (***DS)***: This model allows the selectivity to be dome shape for the main purse seine fishery on yellowfin associated with dolphins and estimates the parameters of the double normal selectivity curve. This will allow the model to fit the length composition data better.

The combination of these hypotheses comprises the reference set of models for the assessment of yellowfin tuna in the EPO (Table 3).

5. RESULTS

5.1. Model diagnostics

5.1.1. Model convergence

All the 48 model runs for yellowfin converged (produced positive definite Hessians matrices), and 28 had small maximum gradients (< 0.001), eight had large maximum gradients (>1) (Table 4). Those with large gradients are TBM.DS and TBM.GRO for all steepness values.

5.1.2. Fit to purse-seine indices of abundance

The RMSE and the negative log-likelihood (NLL) of the purse-seine index of abundance are used to evaluate how well the models fit that data (Table 5, Figure 11). Small values for both RMSE and NLL suggest the assessment fits the data well. The models that best fit the index of abundance were those with the

assumption of density dependence catchability (DDQ), followed by those with the time-block in the middle of the series (TBM) (Figure 11 and Appendix A3). Of those, the models that best fit the index and had less residual patterns (Figure A3) were the ones that estimated growth (DDQ.GRO). Changing the steepness did not improve the fit; within a configuration, the models with different steepness fit the index about the same.

The model configurations that showed fewer residual patterns over time where those that assumed a non-linear relationship between the index of abundance and its vulnerable biomass (DDQ, DDQ.GRO and DDQ.DS) (Figure A3). They were followed by the models with a time-block in the middle of the period (TBM, TBM.GRO, TBM.DS). A seasonal pattern was present in the residuals of all models , with positive residuals in quarters 1 and 2 and negative residuals in quarters 3 and 4 (Figure A4).

5.1.3. Fits to length frequency data

The results of this section focus on the model with steepness equal 1 because the fit of models with the same configuration, except for the steepness assumption, fit the composition data almost identically, with at most one negative log-likelihood unit of difference (Table 6).

For all fisheries and for the survey, the average effective sample size (based on McAllister and Ianelli 1997) is about 5 to 20 times larger than the input adjusted sample size (Table 6). As expected, the survey length frequencies are fit best by models that have blocks in selectivity. The fit of different model configurations to the length frequency composition is very similar for all but three fisheries: F3-OBJ-C_Q14, F18-DEL_C and F19-DEL_*P*. F19-DEL_*P* is the fishery that has the asymptotic selectivity assumption in three of the models, and dome-shape in three models, and blocks of selectivity (asymptotic and dome-shape) in six models. The models that fit this data best are TBM.DS. F18-DEL_C is one of the fisheries with the largest catches during all the assessment time period (Figure 2), and it is the fishery that shows an increase in the average size in recent years (Figure 5B) . These data are fit best by the models that consider that the index is non-linearly related to abundance (DDQ, DDQ.DS and DDQ.GRO). Finally, F3-OBJ-C_Q14 is a floating object fishery that expanded after the mid-1990s and has a skewed length frequency distribution towards larger sizes (Figure 5A).The models that fit these data best are those that estimate growth (DDQ.GRO, GRO, TBE.GRO and TBM.GRO).

Another way to visualize how well the models fit the length composition data is through the empirical selectivity. The two curves should be similar if the model is a good fit to the data. Figure 14 shows this plot for BASE h=1. The fit to most of the length frequency data is similar, except for F19-DEL_*P*, which is fit better by the model that assumed dome-shaped selectivity (Figure 15).

Residual plots are shown for the survey, and fisheries F18-DEL_C and F19-DEL_*P* (Figure A5). The negative residuals after 2000 improve using the models with a time block at that time (TBM, TBM.GRO, TBM.DS). The trend towards positive residuals at the end of the time series is reduced by the models that have a time block at the end (TBE, TBE.GRO and TBE.DS). Finally, the trends towards negative residuals for larger sizes for F19-DEL_*P* is improved by the models that either used dome-shaped selectivity or estimated growth.

5.1.4. Overall fit

The overall fit was assessed using AIC (*Akaike Information Criterion*). Because some models are also fit to the conditional age-at-length, the AIC was computed without this component to make it more comparable among models. The comparison is an approximation, however, since the models that do not use the conditional age-at-length are expected to have better AIC scores than the models that use it; this is because the latter models' fit to the data will still be affected by the conditional age-at length. The models that fit the data best where TBM.DS and any steepness (Table 7).

5.1.5. Integrated model diagnostics

5.1.5.a Age-structured production model and catch-curve analyses

The ASPMs do not show similar trends to their reference models (Figures A6-A8). The ASPM for all models starts from a depleted state, then increases to two to three times the virgin biomass, and stabilizes at a spawning biomass ratio of one in 1990 to 2019. ASPM models estimating recruitment deviations (ASPM-R) show population trajectory trends that are more similar to the reference model but at a much lower spawning biomass ratio. Only the ASPM-R with density depended catchability (DDQ) configurations (DDQ, DDQ.GRO and DDQ.DS) have a positive definite Hessian. Of these, only the DDQ configuration, however, has confidence intervals that overlap with the reference model. These results indicate that information on relative recruitment over time is needed to extract absolute abundance information from the abundance index.

The CCAs are aligned with reference models in several periods (Figure A7). The CCAs, both the one based on the standardized length composition corresponding to the index of abundance and the one based on the fisheries length compositions, show similar results. The beginning of the series for BASE, TBM and GRO, are markedly different between the refence models and the CCAs. The implied index of abundance for the CCAs also shows that for an index to follow the same pattern that the CCAs are inferring, it should be stable along the whole period, but with large interannual variations (Figure A8). This, in addition to the ASPM results above, indicates that the scale of the models is highly influenced by the composition data, but the index of abundance is needed to inform or constrain the temporal variation and trends.

When the trajectories of the CCA and the integrated models are dissimilar, there is an indication of model misspecification (Figure A7). The discrepancies observed in the BASE model around the year 2000 and from 2015 on, are solved by the TBM and TBE configurations. The DDQ reference models have better overall match with the CCAs. This indicates that the length composition data and the index tend to support similar trajectories when either the index is assumed to have a non-linear relationship with the vulnerable biomass or when there is a block of catchability and selectivity for the index and the main fisheries (F18-DEL_C and F19-DEL_*P*).

Some assumptions seem key to solving model misspecification while others have no effect. The dome-shape selectivity assumption and the change in catchability seem key, while estimating growth is not important. The TBE and the DS reference models have better correspondence with the CCAs except in 2001-2007, when the CCA predicts smaller biomass than the reference models, and around 2010 when the CCA predicts larger biomass (Figure A7). The TBM solves the 2001-2007 discrepancies. The models that estimate growth have similar patterns to those with fixed growth (and identical otherwise). This indicates that understanding the changes in fisheries strategies (*e.g.* gear, search behavior, market demands for larger fish) are a central part for assessing the yellowfin tuna population.

5.1.5.b Likelihood profile on R_0

This diagnostic is helpful in determining the relative importance of different data components on the estimates. The likelihood profile on R_0 (in log scale) indicates that the results for all model configurations are driven by the length composition data, except for GRO and TBE.GRO (Figure A9). For TBM, and to some extent for TBM.GRO, although the length composition data is the most influential component, it is not in contradiction with the index of abundance. This means that including a time-block when the index indicates large abundance may resolve some model misspecification.

5.1.5.c Retrospective analyses

The retrospective analyses show the behavior of the models when new data are added. Two model configurations showed instabilities when data from progressive years were removed (DDQ.DS for both

spawning biomass ratios (SBR, ratio of the index of spawning biomass over the virgin spawning biomass) and *F*, and TBM.DS, for F) (Figures A10 and A11). These issues were not resolved even after several attempts to start the model fits from different starting value. The other model configurations that treated the index as one continuous series linearly related to the biomass (BASE, GRO and DS) showed retrospective pattern, both in the spawning biomass ratio and in the fishing mortality. The other models show no important changes in the results by sequentially removing data for the terminal years, indicating that addressing misspecifications in the observation model for the index of abundance, and in the selectivities for the main fleets and survey, improves the robustness of the models.

The previous assessment was not considered reliable for management advice because it was too sensitive to the addition of new abundance index data from the longline fishery (SAC-10-INF-F). These data are no longer used in the assessment and this may be why most models are now not over-sensitive to new data.

5.1.6. Parameter estimates

5.1.6.a Initial conditions

All models estimate that the population starts from a depleted state in 1984 (Spawning biomass ranging from 14% to 72% of unexploited)

5.1.6.b Selectivity

In general, estimated selectivities follow the empirical selectivities well (Figure 12), except for fishery F19-DEL-*P*, for which the fit depends on the model assumptions. Differences in the selectivity of F19-DEL-*P* occurs primarily at large sizes (Figure 13). When assuming asymptotic selectivity, the selectivity at sizes >100 cm is estimated to be 1. In comparison, when assuming dome-shaped selectivity, selectivity is estimated to reach the peak (*i.e.*, 1) at around 100 cm before dropping to a final (at L_2) level of about zero (DS runs). The runs that estimate growth predict no or a very low proportion of large fish (>175cm). The models with a time block at the end (TBE)have better correspondence between the estimated selectivity for smaller sizes and the empirical selectivity . This is because most of the time series has dome-shaped selectivity (Figure 15). Fit is improved by the models that either used dome-shaped selectivity or estimated growth. However, models that estimated growth predicted no fish in some size classes with fish of those sizes in the observations (Figure 14). The TBM models estimated differences in selectivity for small fish in the index of abundance and for large fish in the fisheries. The TBE models estimated a shift towards larger sized in the selectivity curve of the F19- DEL-*P* fishery after 2015, as expected (Figure 14).

5.1.6.c Catchability and density dependence

The catchability estimates for the index of abundance at the end of the time-series ranged from 86% to 91% of the catchability value for the earlier period (Table 8). The catchability estimates for the index of abundance in the block in the middle of the time-series are 164% to 182% of the catchability for the rest of the time period, almost doubling the catchability during that period. The density-dependent parameter ranged from 1.7 to 2.1.

5.1.6.d Growth

The estimated growth curves differ in two main ways from the fixed values (Figure 15): L_1 is about 21 cm instead of 18.4 cm and L_2 is smaller than the fixed value. The estimates of L_2 range from 149.4 to 161.9 cm. Because L_2 is the average size of the oldest fish, and given the assumed CV=0.075 of variation of length at age, a fish of age 29 quarters as large as 182.4 or more might still be found in the population, but with smaller probability (about 16 fish in 10,000 would be that large or larger). Another implication of the estimated growth function is that fish stop growing at about 4 years of age.

5.2. Stock assessment results

5.2.1. Recruitment

Time series of estimated quarterly age-zero recruits are shown in Figure 16A, annual recruits in Figure 16B.

The recruitment estimates are not sensitive to the value of steepness. All models estimate an initial period of above average recruitment, which culminated with the largest recruitment in 1999 and lasted to about 2000, followed by lower than average recruitments. This pattern follows the general trend shown in the index of abundance. All models have an increase in the point estimate of recruitment in the last year, but with a large confidence interval. This large confidence interval is expected since there is not much information in the data about recruitment in the most recent year.

The general patterns vary in their magnitude for the different configurations. For example, the large 1999 recruitment is much larger in the DS model than in the TBM.DS model, where most of the increase in the index is attributed to a change in catchability, rather than the result of a very strong recruitment. In the TBM.DS model, the 1999 recruitment is comparable to the 1993 one. The DDQ models have less pronounced difference between the high and the low recruitment periods, as much of the differences are absorbed in the non-linear relationship between the index and abundance. The DS models have the opposite pattern: recruitment in the high recruitment period is much larger than in the low recruitment period. Models that estimate growth have similar point estimates of recruitment as their fixed counterpart but have an increase in the uncertainty of the estimates, as expected. Assuming a dome-shape selectivity for fleet F19-DEL_*P* also increased the uncertainty of the recruitment estimates, especially

Similar to the recruitment results, spawning biomass estimates are not sensitive to steepness (Figure 17). However, they differ in a key point: the estimate of equilibrium virgin biomass. This translates to differences among the steepness runs in the spawning biomass ratios (Figure 18). The models with steepness equal to 1 have the least depleted series and those with steepness of 0.7 have the most depleted series. All the point estimates of the trajectories with different steepness, however, are within the confidence intervals of each other.

All biomass trajectories have declining trends, but they vary in the magnitude of the declines. In one extreme, those with the most pronounced declines assume a time-block in the middle of the timeseries (2001-2003) of the index of abundance, which isolates the period of the sudden increase in the abundance index with its own catchability, and the TBM.DS has the largest biomass decline. The models that assume that that increase in the index of abundance in 2001-2003 are real, and not a sudden change in catchability, estimate less declines over time (*e.g.* BASE, GRO, DS, TBE). At the other extreme, the models that assume a non-linear relationship between the index and the biomass (DDQ, DDQ.GRO and DDQ.DS) estimate the least declines. All models show the lowest SBR in mid-2016 and an increasing trend afterwards.

5.2.2. Fishing mortality (*F*)

Regarding fishing mortality, similarities and contrasts among models are also apparent (Figure 18). The main similarities are in the relative magnitude of fishing mortality (*F*) between age classes. All models have the highest *F* on fish aged 21+ quarters (5.25 years-old and older), followed by fish aged 11 to 20 quarters (2.75-5 years). The lowest fishing mortality at age is on the youngest fish and is about the same for all models. All models estimate an increase in fishing mortality for the two oldest age classes over time.

The main difference among models are in the magnitude of *F* for the oldest age classes. The models with fixed growth have the highest *F* for age 21+ fish, when comparing with those that estimate growth or assume dome-shape selectivity. This is because the fixed growth models assume that older fish have larger sizes (L_2=182.6 cm), and since these sizes are rare in the observations, that means that the fishing mortality

must be high. The models that estimate growth explain the scarcity of those large size fish in the observations by decreasing the average size of the older fish (L_2 from 149.4 to 161.9 cm) making the proportion of large fish smaller (given the CV of length at age is 7.5%), as seen in the data. The models that assume dome-shape selectivity assume that the reason there are no large fish in the data is because those sizes are not vulnerable to the purse-seine fisheries that operate north of 5°N (the fisheries that provide the length frequency data used to fit the models). The models that estimate a time block in selectivity of both the index and the F19-DEL_*P* fishery, from 2015 on (TBE,TBE.GRO and TBE.DS), isolating the period of large increase in sizes with a different selectivity, assume dome-shape selectivity before 2015, thus explain the lack of large size in the past by assuming those fish where not vulnerable to the purse-seine fishery before 2015. It is intriguing that models with a time-block in the middle (TBM, TBM.GRO and TBM.DS) estimate the same low fishing mortality as the TBE ones. By using a time block for the index catchability (2001-2003) and selectivity (2002-2007), the TBM model estimate a biomass trajectory analogous to the TBE models, given that the catches are fixed, that translates into similar fishing mortalities.

5.3. Fisheries impacts

This is analysis compares the impact on the spawning biomass of fisheries that have different selectivities (Wang *et al* 2009). The impact for each type of fishery was estimated by projecting the population without their catches and obtaining the resulting spawning biomass (index of spawning output in this assessment). The increased spawning biomass in the absence of the catches of those fisheries in relation to the current spawning biomass indicates the impact of those fisheries.

All models estimate similar impacts of the different types of fisheries (Figure 20). The longline and the sorted discards have the smallest impact while the largest impact over most of the period is by purse-seine associated with dolphins. The unassociated fisheries had the second largest impact in early years. Since about the nineties, the impact of the floating object fisheries started to be important. The impact of these fisheries surpassed that of unassociated fisheries around 2008 and ten years later, in 2018, it surpassed the impact of the purse-seine fisheries associated with dolphins. In 2019 the fisheries with the largest impact where the floating object fisheries

5.4. Comparison with the previous assessment

One of the main differences between this assessment and the previous one is on its use of the data available. Previous assessments, including the SAC10 assessment, were fit to 5 indices of abundance, one from the longline fishery and four from the purse-seine fisheries. The longline index was based on standardized CPUE from the Japanese fleet. The purse-seine indices were nominal CPUEs and were limited to certain areas of the EPO. The purse-seine and the longline indices had inconsistencies that were considered a major issue for the previous assessments. A new spatio-temporal modeling framework was developed and applied to the CPUE data to create new indices, but the inconsistencies were not resolved. Standardized length frequencies suggest that the two indices may be indexing different groups of fish. The prominent index peak in 2001-2003 that seems to occur earlier in the longline index and later in the purse-seine index, opposite to what was expected given the growth and selectivity assumptions of the model, was due mainly to the 1998 cohort (of an important El Niño year) in longline fisheries and to the 1999 cohort (of an equally important La Niña year) in purse-seine fisheries. Why these indices tracked those two cohorts differently is still an unresolved issue for future research. Also, how or if other cohorts, of smaller magnitude, may be subject to the same phenomenon, it is unknow. One of the hypotheses is spatial heterogeneity, which is somewhat addressed in the current model that is fit to a purse-seine index of abundance for the EPO north of 5°N and also to the length composition data from the purse-seine fisheries that fish north of 5N, but not fisheries to the south.

Results of the current assessment and the previous differ in the uncertainty of the estimates (Figure A12). The uncertainty in the SAC10 model was very small, because there were limited combinations of parameter values that allowed the model to fit the contradictory information from the indices, and from the larger data weighting of length composition data relative to new models, which implement the Francis method of composition data reweighting. The confidence intervals for any model in the current assessment are much wider than the SAC10 model.

Another important difference between the SAC10 model and the current assessment are the fishery definitions and the assumed selectivity for each fishery. The fisheries in the current assessment were defined using a regression tree analysis that maximized the differences in size composition among fisheries and minimize the difference within fisheries (in space and time). Then, splines were used to best characterize the selectivity for each fishery. Splines allowed more flexibility in the shapes of the selectivity functions than those used for the SAC10 model (double-normal functions). In the current assessment the mortality at size was better characterized, which changed the estimated fishing mortality at age. The SAC10 model estimated larger fishing mortality at age for ages 10 to 21 (Figure 3 in SAC-10-07), while this assessment estimates that the fishing mortality for fish age 21+ may be as large or larger (Figure 19) . All models have a lower fishing mortality for fish younger than 10 quarters, similarly to SAC10, but even lower than SAC10, due to a larger number of recruits being estimated in all models. This is also related to selectivity functions with a narrower range of lengths in the floating object fisheries, corresponding to better fits to the length frequencies, which translate in some ages not being fully selected to the floating objects.

The relative impact of different fisheries is similar from this assessment (Figure 20) and SAC10 (Figure 4 in SAC-10-07) as well as the tendency of increase in F. The tendency for an increase in the impact of the floating object fisheries and the decrease of impact of the unassociated fisheries shown in previous assessments, is also estimated for all models of the reference set. The overall fishing mortality has increased in recent years, similar to what was estimated by the SAC10 assessment.

6. STOCK STATUS

The stock status of yellowfin tuna in the EPO is assessed by considering calculations based on the spawning biomass and the maximum sustainable yield (MSY). Maintaining tuna stocks at levels capable of producing MSY is the management objective specified by the IATTC Antigua Convention.

6.1. Definition of reference points

Resolution C-16-02 defines target and limit reference points, in terms of biomass and fishing mortality. Based on the resolution, the definitions of those reference points are listed below together with the method used to compute them in this document.

6.1.1. Limit reference points

The spawning biomass limit reference point (S_{LIMIT}) is the spawning biomass threshold that should be avoided because being more depleted could endanger the sustainability of the stock. The S_{LIMIT} adopted as interim by the IATTC (in the 87th meeting) is the spawning biomass that produces half of the virgin recruitment given that the stock-recruitment relationship follows the Beverton-Holt function with a steepness of 0.75. This spawning biomass is equal to 0.077 of the equilibrium virgin spawning biomass (Maunder and Deriso 2014 – SAC-05-14). The resolution further established an interim harvest control rule that requires action be taken if the probability of the spawning biomass at the beginning of 2020 ($S_{current}$) being below S_{LIMIT} is greater than 10%. Thus, for the purpose of providing management advice, $S_{current}/S_{LIMIT}$ and the probability of this ratio being smaller than 1 (by assuming the probability distribution function for the ratio is normal) are included in the management table.

The limit reference point of fishing mortality (F_{LIMIT}) is the fishing mortality threshold that should be avoided because fishing harder could endanger the sustainability of the stock. The F_{LIMIT} adopted as interim by the IATTC (in its 87th meeting) is the fishing mortality rate that, under equilibrium conditions, maintains the spawning population level at S_{LIMIT}. The interim harvest control rule requires action to be taken if the probability of the average fishing mortality in 2017-2019 ($F_{current}$) being above F_{LIMIT} is greater than 10%. Thus, to provide management advice, $F_{current}/F_{LIMIT}$ and the probability of this ratio being larger than 1 (by assuming the probability distribution function for the ratio is normal) are included in the management table.

6.1.2. Target reference points

The target reference point of spawning biomass is the level of spawning biomass that should be achieved and maintained. The spawning biomass that produces the MSY (S_{MSY}) was adopted by the IATTC (in the 87th meeting) as the interim target reference point for tropical tunas in the EPO. The resolution mentions that when dictating rebuilding targets, actions are required to allow for at least 50% chance of rebuilding the spawning biomass to the dynamic MSY level (S_{MSY_d}) within 5 years or two generations. Here, S_{MSY_d} is derived by projecting the population into the future under historical recruitment and a fishing mortality rate that produces MSY (F_{MSY}). The value of S_{MSY_d} used to compute reference points for yellowfin is the mean projected spawning biomass for the last four projection quarters. To provide management advice, $S_{current}/S_{MSY\_d}$ and the probability of this ratio is smaller than 1 (by assuming CV is equal to that of $F_{current}/F_{MSY}$) are included in the management table.

The target reference point of fishing mortality is the level of fishing mortality that should be achieved and maintained. F_{MSY} was adopted by the 87th meeting of the IATTC as the interim target reference point for tropical tunas in the EPO. Thus, to provide management advice, $F_{current}/F_{MSY}$ and the probability of this ratio is larger than 1 (by assuming the probability distribution function for the ratio is normal) are included in the management table. Also, the inverse of $F_{current}/F_{MSY}$, *F* multiplier, is also included in the management table.

In the Kobe plot (Figure 21), the time series of S_{MSY_d} is computed based on two approximations: (1) $S_{MSY\_d1} = S_{0\_d}\,(S_{MSY}/S_0)$, where S_{0_d} is the dynamic spawning biomass with the absence of fishing and S_{MSY}/S_0 is the depletion level that, under equilibrium, produces the maximum sustainable yield; (2) S_{MSY_d2}, which is derived by projecting the population for the future under $F=F_{MSY}$ and historical recruitment. The two approximations are weighted as follows to obtain the trajectory of $S_{MSY\_d}(t)$ in the Kobe plot:

$$S_{MSY\_d}(t) = p(t)\, S_{MSY\_d1}(t) + (1- p(t))\, S_{MSY\_d2}(t) \quad \text{Equation 8)}$$

where *p* increases linearly as a function of year (t) from 0 in the start year to 1 in the end year.

The dynamic MSY (MSY_d) in the management table is also derived by projecting the population into the future under historical recruitment and F_{MSY}. We assume that MSY_d is the sum of the projected total fishery catches for the last four projection quarters.

6.2. Estimates of stock status

According to the 48 reference models, the point estimate for the spawning biomass at the beginning of 2020 ranged from 49% to 219% of the dynamic spawning biomass at MSY (Table 9, Figure 21). The probability that the spawning biomass at the beginning of 2020 is lower than the dynamic MSY level is 50% or less for 13 of the 48 models. The point estimate of the 2017-2019 fishing mortality ranged from 40% to 168% of the fishing mortality at MSY. The probability that the fishing mortality in 2017-2019 was higher than the MSY level is 50% or more for 14 of the 48 models.

The point estimate for the spawning biomass at the beginning of 2020 ranged from 145% to 345% of the

S_{LIMIT}. The probability that the spawning biomass at the beginning of 2020 is below S_{LIMIT} ranges from 0 to 2%. The point estimate of the 2017-2019 fishing mortality ranged from 22% to 65% of the F_{LIMIT}. The probability that the fishing mortality in 2017-2019 was higher than F_{LIMIT} was estimated to be zero for all models.

Every reference model suggests that a lower steepness value corresponds to more pessimistic estimates of stock status: lower spawning biomass relative to the reference points and higher fishing mortality relative to reference points. However, regardless of what value is assumed for steepness, all models that assume either fixed growth, a linear relationship between the index of abundance and the vulnerable biomass or no changes in selectivity and asymptotic selectivity for the purse-seine fishery that catches the largest fish, estimate the stock to be overfished ($S<S_{MSY\_d}$) and that overfishing is occurring ($F>F_{MSY}$). All models that assume dome-shape selectivity for the purse-seine fishery that catches the largest fish, estimate the stock as not overfished ($S>S_{MSY\_d}$) and that overfishing is not occurring ($F<F_{MSY}$). There is considerable uncertainty associated with those estimates (Figure 21); several models that are in the green quadrant of the Kobe plot (Figure 21) have confidence intervals that include the yellow and red quadrants, implying that those models also provide some support for the hypotheses that the stock is overfished and/or that overfishing is occurring.

The results of all this models are used in a risk analysis (SAC-11-08) to evaluate the probability of exceeding the reference points specified in the harvest control rule.

7. FUTURE DIRECTIONS

Recommendations of the external review panel, as well as lessons learnt in this benchmark assessment will be taken into account in future work. Specifically, the staff plans to focus on:

Collection of new and updated information

- Continue its collection and analysis of purse-seine data (catch, effort, and size-composition)
- Continue collaborative work with longline CPCs
- Continue tagging and biology studies and analyses

Refinements to the assessment model and methods

- Address uncertainty in spatial/stock structure
- Continue research on CPUE and length frequency standardization methods
- Work with purse-seine CPCs to understand changes in fishing strategies to inform selectivity modelling
- Continue exploring uncertainty in growth and selectivity
- Explore uncertainty in natural mortality
- Explore different stock assessment time spans, initial conditions and types of models (monthly/weekly models, depletion models)
- Explore other integrated model diagnostics

8. ACKNOWLEDGEMENTS

Many IATTC and member country staff provided data for the assessment. IATTC staff members, and member country scientists provided advice on the stock assessment, fisheries, and biology of yellowfin tuna. Nick Webb provided editorial assistance and Christine Patnode aided on the figures.

9. REFERENCES

Aires-Da-Silva, A., and Maunder, M. N. 2009. Status of yellowfin tuna in the eastern Pacific Ocean in 2007 and outlook for the future. IATTC Stock Assessment Report 9: 3-94

Bayliff, W.H. 1979. Migrations of yellowfin tuna in the eastern Pacific Ocean as determined from tagging experiments initiated during 1968-1974. Inter-Amer. Trop. Tuna Comm., Bull. 17: 445-506.

Bayliff, W.H. 1979. Migrations of yellowfin tuna in the eastern Pacific Ocean as determined from tagging experiments initiated during 1968-1974. Inter-Amer. Trop. Tuna Comm., Bull. 17: 445-506.

Bayliff, W.H. 1984. Migrations of yellowfin and skipjack tuna released in the central portion of the eastern Pacific Ocean, as determined by tagging experiments. Inter-Amer. Trop. Tuna Comm., Intern. Rep. 18: 107 *p*

Beverton, R.J., and Holt, S.J. 1957. On the dynamics of exploited fish populations. Fisheries Investigation Series 2, volume 19, UK Ministry of Agriculture. Fisheries, and Food, London, UK.

Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., and Piner, K.R. 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research **192**: 28-40.

Estes, D. H. 1983. Shio-Japanese pioneers in San Diego's Fishery. In: Boat and Ship building in San Diego; 11th Annual Cabrillo Festival Historic Seminar. San Diego California, Cabrillo Historical Association. 1: 25-43

Fink, B.D. and W.H. Bayliff. 1970. Migrations of yellowfin and skipjack tuna in the eastern Pacific Ocean as determined by tagging experiments, 1952-1964. Inter-Amer. Trop. Tuna Comm., Bull. 15: 1-227.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences **68**(6): 1124-1138.

Grewe, *P*., Hampton, J. 1998. An assessment of bigeye (*Thunnus obesus*) population structure in the Pacific Ocean, based on mitochondrial DNA and DNA microsatellite analysis. CSIRO Marine Research, Hobart, Australia.

Hampton, J. 2000. Natural mortality rates in tropical tunas: size really does matter. Canadian Journal of Fisheries and Aquatic Sciences 57(5): 1002-1010.

Hampton, J.A., and Fournier, D.A. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshwater Res **52**, 937–963

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, Vert-Pre, K.A, Whitten,A.R., Punt, A.E. 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models ICES Journal of Marine Science 72(1), 99–110. doi:10.1093/icesjms/fsu19

IATTC 2010. The IATTC program for in-port sampling of tuna catches. IATTC Document SAC-01-11. http://www.iattc.org/Meetings/Meetings2010/Aug/_English/SAC-01-11-Port-sampling-program.pdf

Itano, D. G., 2000. The reproductive biology of yellowfin tuna (*Thunnus albacares*) in Hawaiian waters and the western tropical Pacific Ocean: project summary. SOEST 00–01, JIMAR Contribution 00–328. Pelagic Fisheries Research Program, JIMAR, University of Hawaii.

Lee, H. H., Maunder, M.N., Piner, K.R., and Methot, R.D. 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fisheries Research **125**: 254-261.

Lee, *H.H*., Piner, K.R. , Methot, R.D., and Maunder, M.N. 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: an example using blue marlin in the Pacific Ocean. Fish. Res. 158: 138-146.

Lennert-Cody, C.E., and Tomlinson, *P*.K. 2010. Evaluation of aspects of the IATTC port sampling design and estimation procedures for tuna catches. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 10: 279-309.

Lennert-Cody, C.E., Minami, M., Tomlinson, *P*.K., Maunder, M.N. 2010. Exploratory analysis of spatial-temporal patterns in length-frequency data: An example of distributional regression trees. Fisheries Research 102: 323-326

Lennert-Cody, C.E., Maunder, M.N., Tomlinson, *P*.K., Aires-da-Silva, A., Pérez, A. 2012. Progress report on the development of postratified estimators of total catch for the purse-seine fishery port-sampling data. IATTC Document SAC-03-10. http://www.iattc.org/Meetings/Meetings2012/May/_English/SAC-03-10-Post-stratified-estimators.pdf

Matsumoto, T. and Bayliff, W.H. 2008. A review of the Japanese longline fishery for tunas and billfishes in the eastern Pacific Ocean, 1998-2003. Inter-American Tropical Tuna Commission, Bulletin, 24 (1): 1-187.

Maunder, M.N. and G.M. Watters. 2001. Status of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Stock Assess. Rep. 1: 5-86.

Maunder, M.N. and Watters, G.M. 2003. A-SCALA: an age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern Pacific Ocean. IATTC Bull. 22: 433-582.

Maunder, M.N., and Aires-da-Silva, A. 2012. A review and evaluation of natural mortality for the assessment and management of yellowfin tuna in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm. Document YFT-01-07.

Maunder, M.N., and R.B. Deriso. 2014. Proposal for biomass and fishing mortality limit reference points based on reduction in recruitment. Inter-Amer. Trop. Tuna Comm. 5th Scient. Adv. Com. Meeting. SAC-05-14.

Maunder, M.N, Xu, H., Lennert-Cody, C.E., Valero, J.L, Aires-da-Silva, A., Minte-Vera, C. 2020a Implementing reference point-based fishery harvest control rules within a probabilistic framework that considers multiple hypotheses. Inter-Amer. Trop. Tuna Comm., 11th Scient. Adv. Com. Meeting. SAC-09. SAC-11-INF-F

Maunder, M.N., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., Lee, H.H., Kai, M., Chang, S.-K., and Kitakado, T. 2020b. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fisheries Research **229**: 105594

McAllister, M.K. and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sc **54**: 284–300.

Methot, R.D., Taylor, I.G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. **68**: 1744–1760

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research **142**: 86-99.

Methot, R.D., and Wetzel, C.R., Taylor, I.G., and Doering, K. 2020. Stock Synthesis User Manual Version 3.30.15. NOAA Fisheries Seattle, WA. NOAA. Processed Report NMFS-NWFSC-PR-2020-05

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science: Journal du Conseil **56**(4): 473-488.

Nishikawa, Y., Honma, M., Ueyanagi, S., Kikawa, S. 1985. Average distribution of larvae of oceanic species of scombrid fishes, 1956–1981. Far Seas Fisheries Research Laboratories S Series 12, 1–99.

Pecoraro, C., Babbicci, M., France, R., Rico, C., Papetti, C., Chassot, E., Bodin, N., Cariani, A., Bargelloni, L., Tinti, F. 2018. The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. Scientific Reports 8, 13890.

Richards, F. 1959. A flexible growth function for empirical use. Journal of experimental Botany **10**(2): 290-301.

Schaefer, K., Fuller, D., Hampton, J., Caillot, S., Leroy, B., Itano, D. 2015. Movements, dispersion and mixing of bigeye tuna (*Thunnus obesus*) tagged and release in the equatorial Central Pacific, with conventional and archival tags. Fisheries Research 161, 336–355.

Schaefer, K.M. 1998. Reproductive biology of yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. Inter-Amer. Trop. Tuna Comm., Bull. 21: 203-272.

Schaefer, K.M. 2009. Stock structure of bigeye, yellowfin, and skipjack tunas in the eastern Pacific Ocean. IATTC Stock Assessment Report, 9, Status of the tuna and billfish stocks in 2007: 203-221. http://www.iattc.org/PDFFiles2/StockAssessmentReports/SAR9-Stock-Structure-SPN.pdf

Schaefer, K.M., Fuller, D.W., and Block, B.A. 2007. Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the northeastern Pacific Ocean, ascertained through archival tag data. Mar. Biol., 105: 503-525.

Schaefer, K.M., Fuller, D.W., and Block, B.A. 2011. Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the Pacific Ocean off Baja California, Mexico, determined from archival tag data analyses, including Kalman filtering. Fish. Res. 112, 22-37.

Schaefer, K.M., Fuller, D.W., and Aldana, G. 2014. Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in waters surrounding the Revillagigedo Islands Archipelago Biosphere Reserve, Mexico. Fish. Ocean. 23: 65-82.

Schnute, J. 1981. A versatile growth model with statistically stable parameters. Canadian Journal of Fisheries and Aquatic Sciences **38**(9): 1128-1140.

Scott, M.D., Lennert-Cody, C.E., Gerrodette, T., Skaug, H.J., Minte-Vera, C.V., Hofmeister, J., Barlow, J., Chivers, S.J., Danil, K., Duffy, L.M., Olson, R.J., Hohn, A.A., Fiedler, *P*.C., Ballance, L.T., Forney, K.A., 2016. Data available for assessing dolphin population status in the eastern tropical Pacific Ocean. Workshop on Methods for Monitoring the Status of Eastern Tropical Pacific Ocean Dolphin Populations: DEK-01

Shimada, B.M., and Schaefer, M. B.(1956 *A study of changes in fishing effort, abundance, and yield for yellowfin and skipjack tuna in the Eastern Tropical Pacific Ocean.* Inter-Amer. Trop. Tuna Comm Bulletin, 1(7), pp. 347-469.

Suter, J.M. 2010. An evaluation of the area stratification used for sampling tunas in the eastern Pacific Ocean and implications for estimating total annual catches. IATTC Special Report 18.

Thorson, J.T., and Barnett, L.A.K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science **74**(5): 1311-1321.

Thorson, J.T., and Haltuch, M.A. 2018. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences (999): 1-14.

Tomlinson, *P*.K. 2002. Progress on sampling the eastern Pacific Ocean tuna catch for species composition and length-frequency distributions. Inter-Amer. Trop. Tuna Comm, Stock Assess. Rep. 2: 339-365.

Vogel 2014 http://www.iattc.org/Meetings/Meetings2014/May/_English/SAC-05-06-Fishery-in-the-EPO-2013-PRES.pdf

Wang, S.-*P*., Maunder, M.N., Aires-da-Silva, A., Bayliff, W.H. 2009.Evaluating fishery impacts: Application to bigeye tuna (*Thunnus obesus*) in the eastern Pacific Ocean. Fisheries Research **99**:106-111.

Wang, S. *P*., Maunder, M.N., Piner, K.R., Aires-da-Silva, A., and Lee, H.H.. 2014. Evaluation of virgin recruitment profiling as a diagnostic for selectivity curve structure in integrated stock assessment models. Fish. Res. 158: 158-164.

Wexler, J. B., D. Margulies, S. Masuma, N. Tezuka, K. Teruya, M. Oka, M. Kanematsu, and H. Nikaido. 2001. Age validation and growth of yellowfin tuna, *Thunnus albacares*, larvae reared in the laboratory. Inter-Am. Trop. Tuna Comm. Bull. 22:52−91.

Wild, A., and Foreman, T. J. 1980.The relationship between otolith increments and time for yellowfin and skipjack tuna marked with tetracycline. Inter-Amer. Trop. Tuna Comm. Bull. 17: 507-560.

Wild, A., Wexler, J. B., and Foreman, T. J. 1995.Extended studies of increment deposition rates in otoliths of yellowfin and skipjack tunas. Bull. Mar. Sci. 57: 555-562.

Wild, A. 1986. Growth of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean based on otolith increments. Inter-Amer. Trop. Tuna Comm. Bull. 18: 421-482.

Xu, H., Lennert-Cody, C.E., Maunder, M.N., and Minte-Vera, C.V. 2019. Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. Fish. Res. 213: 121-131.

Yamanaka, K. L. 1990. Estimates of age, growth and spawning of yellowfin tuna (*Thunnus albacares*) in the Philippines as determined from examination of increments on sagittal otoliths Tuna Develop. Mgmt. Prog., Coll. Vol. Work. Doc. 3: 27-39.

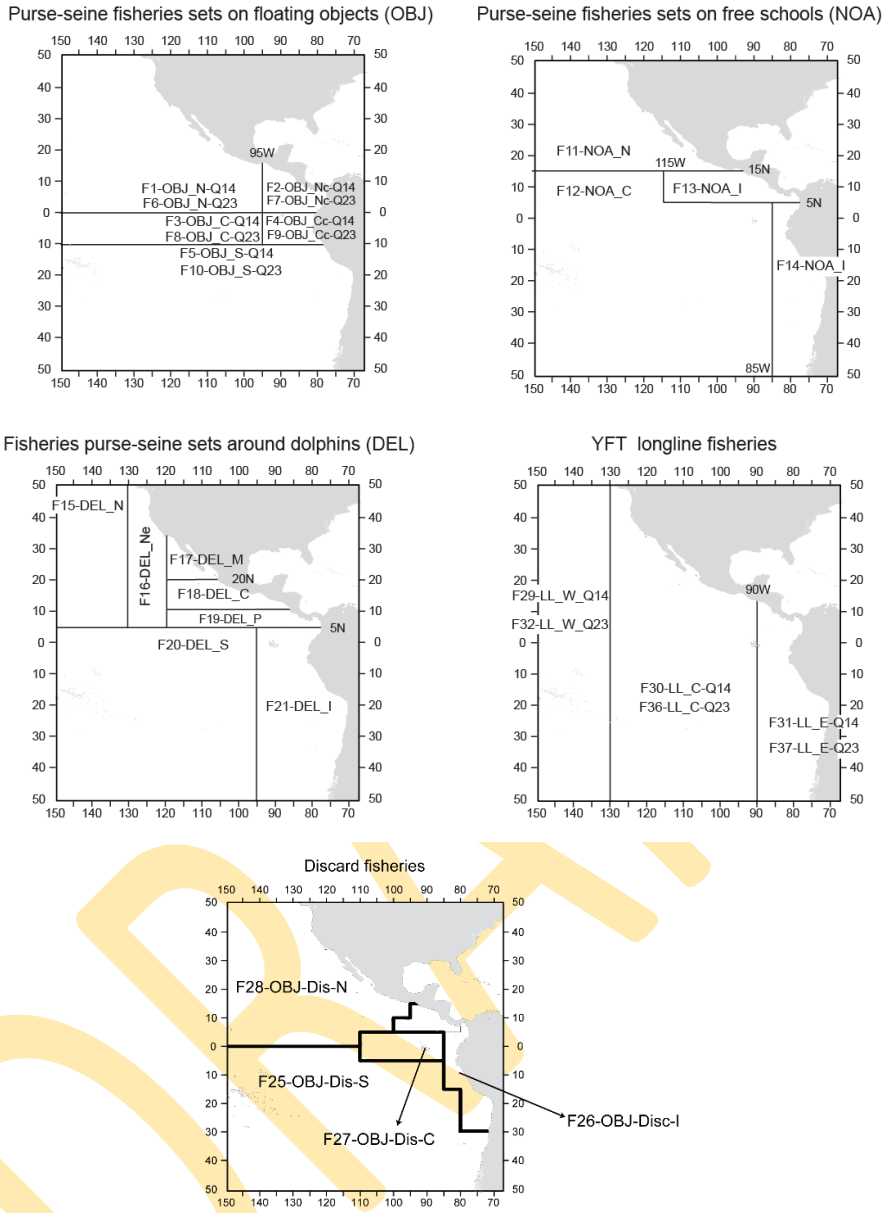

FIGURE 1. Areas corresponding to the fishery definitions used in the stock assessment of yellowfin tuna in the EPO. The lines indicate the boundaries of each fishery, and the alpha-numeric labels the corresponding fishery names. The fisheries are described in Table 1.

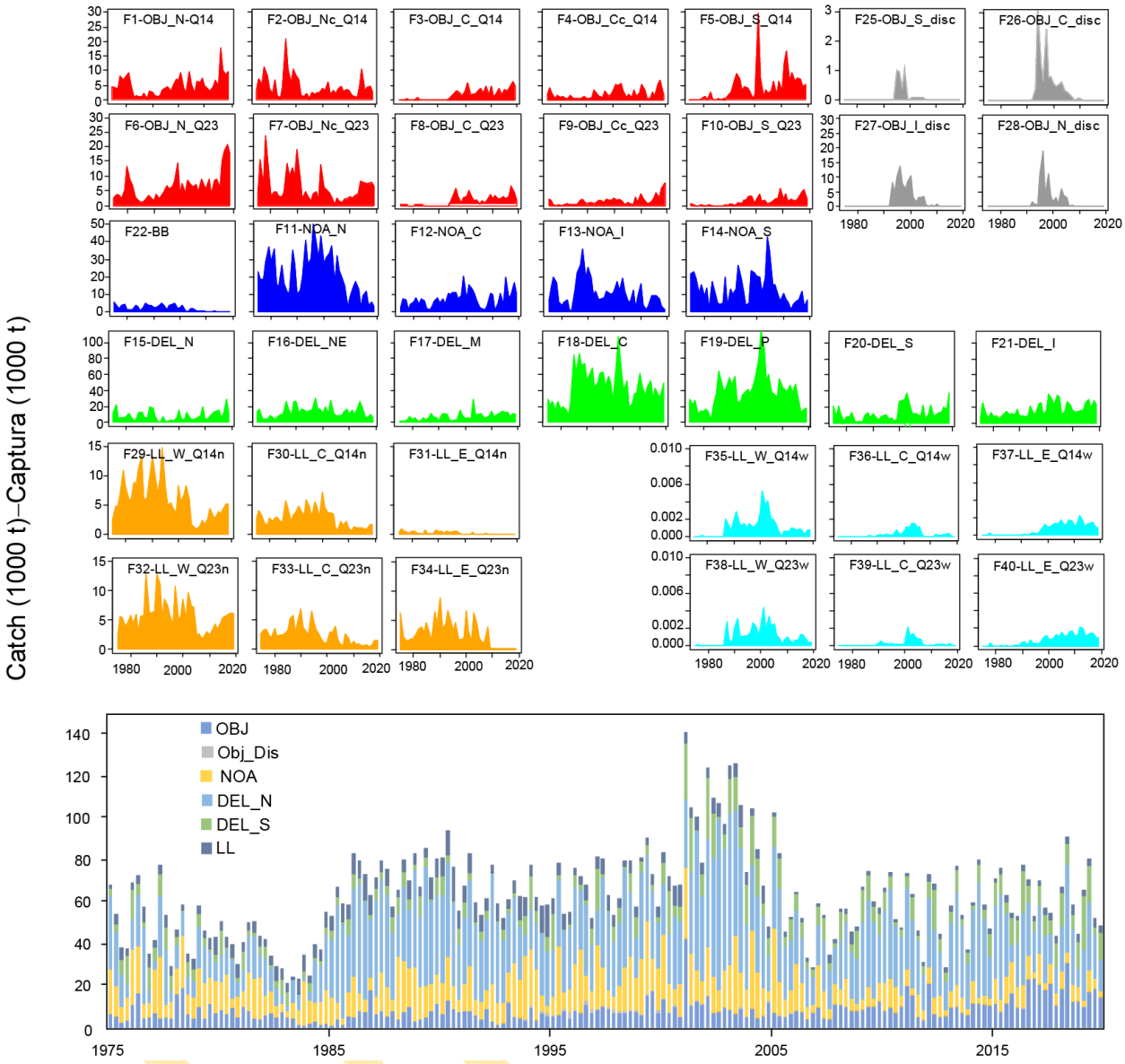

FIGURE 2. (**Top**) Annual catches, by fishery (Table 1), for the stock assessment of yellowfin tuna in the EPO. The panels with different colors have different vertical scales. (**Bottom**) Quarterly catches by gear. Although all the catches are displayed as weights, the stock assessment model uses catches in numbers of fish for longline fisheries F29 to F35. Catches in weight for the longline fisheries F29 to F35 are estimated internally by Stock Synthesis by multiplying the catches in numbers of fish by estimates of the average weights.

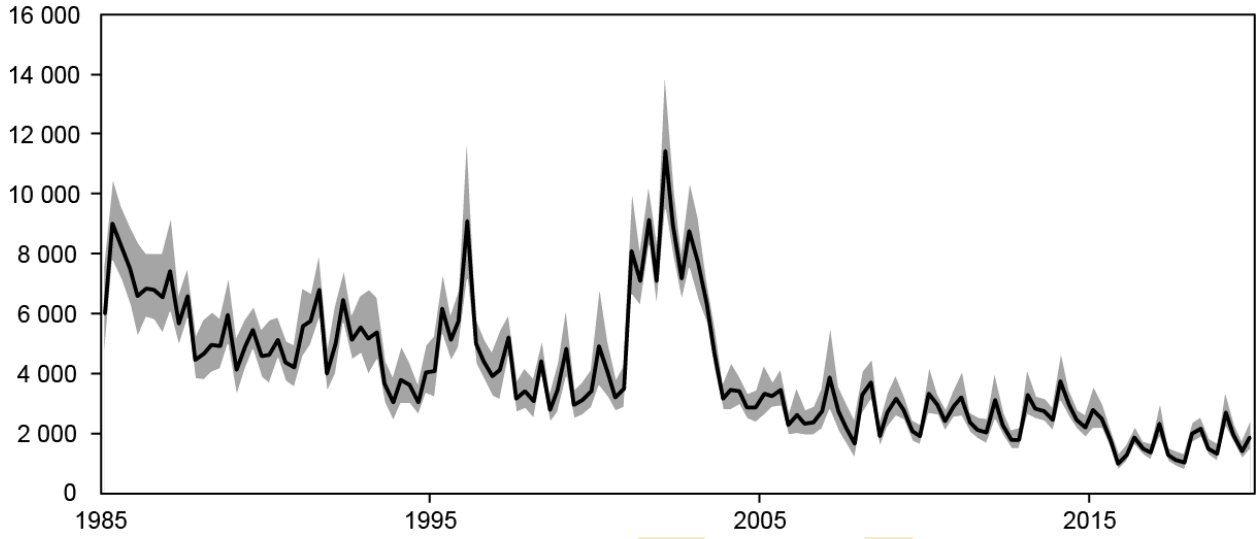

FIGURE 3. The standardized dolphin purse-seine index of abundance (black line) and the associated 95% confidence interval (gray shading) for the stock assessment of yellowfin tuna in the EPO.

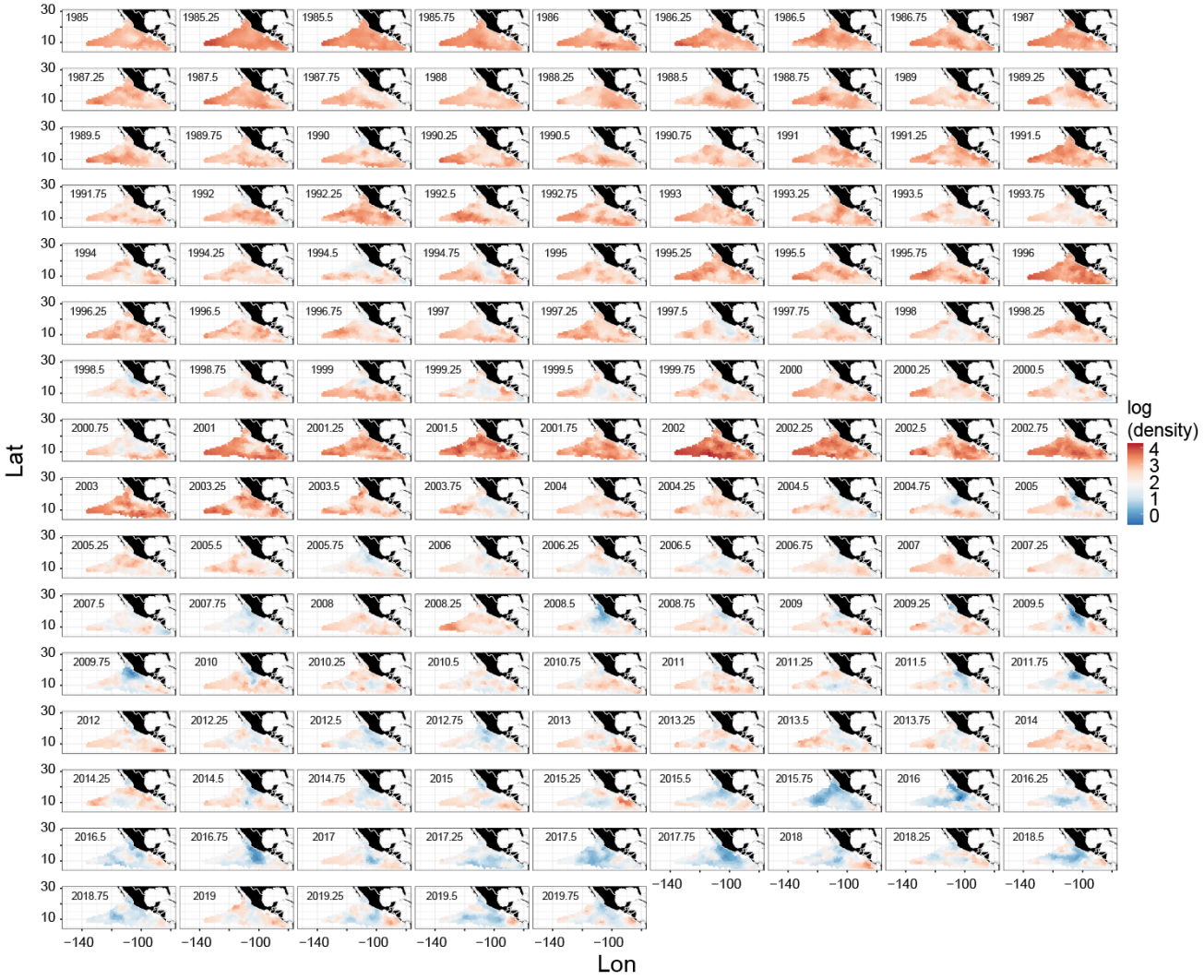

FIGURE 4. Spatiotemporal distribution of the log(density) predicted by the delta-lognormal VAST model for the stock assessment of yellowfin tuna in the EPO.

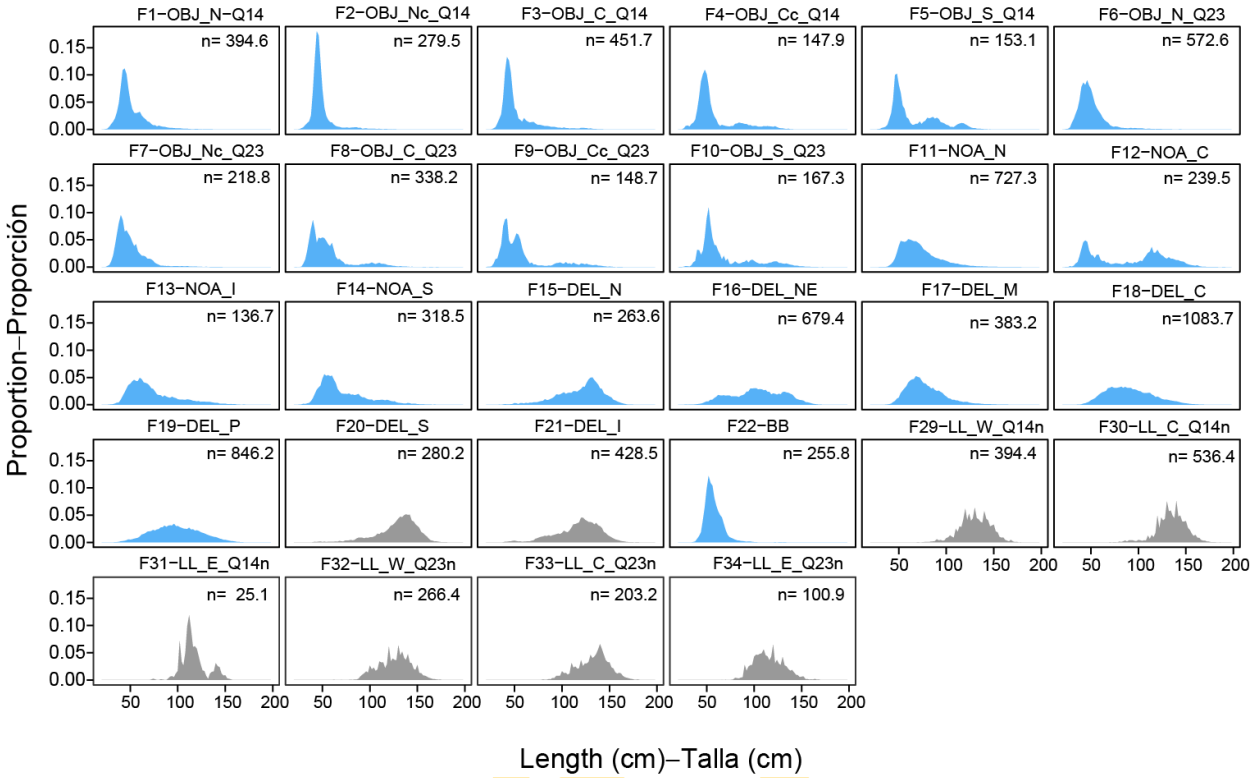

FIGURE 5A. Weighted average observed yellowfin tuna length composition, by fisheries and quarters; *n* is the total adjusted sample size, adjusted by the weight given to the data in the models. In blue: data fit by the reference models; grey: data not fit.

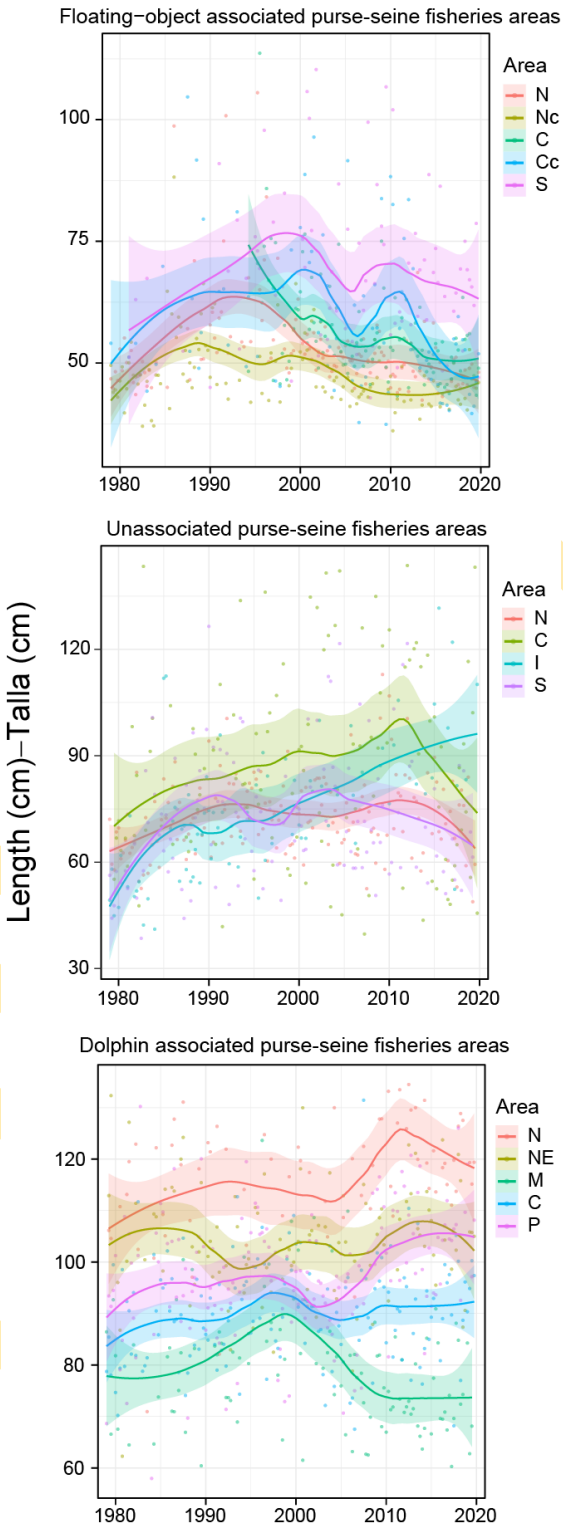

FIGURE 5B. Average length by area over time smoothed with a *lowess* smoother (s=0.5)

.

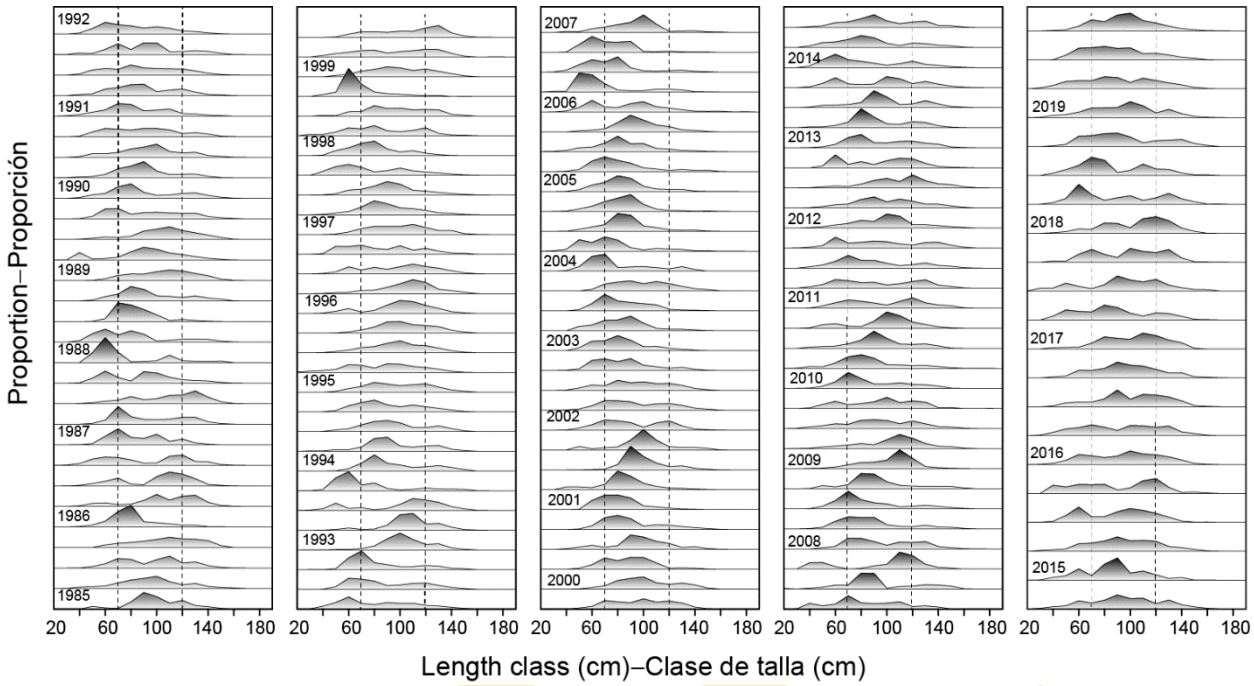

FIGURE 6. Standardized length composition for yellowfin tuna in the EPO, by year for 1985 to 2019. The dashed vertical lines are at 70 and 120 cm.

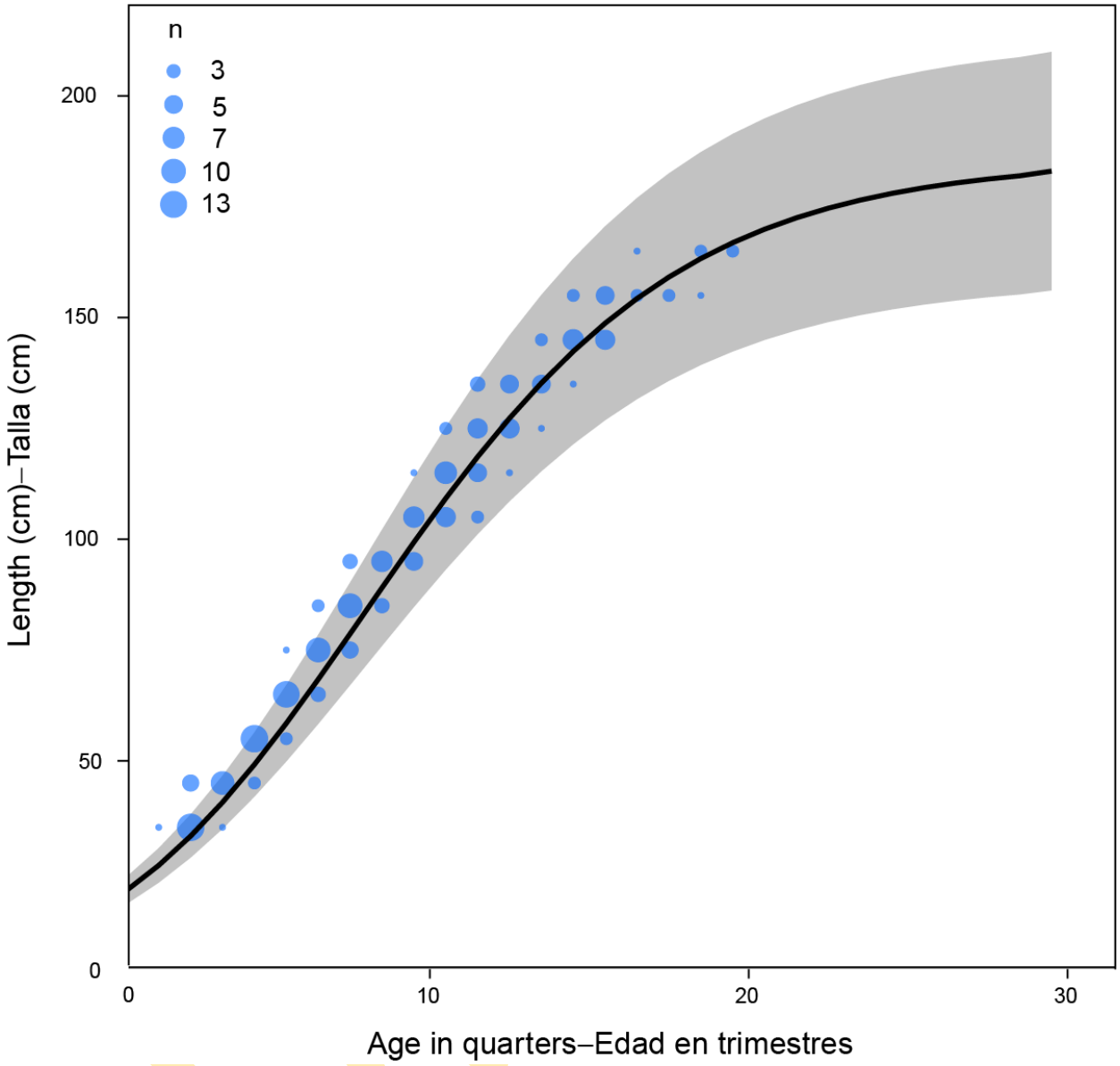

FIGURE 7. Age conditional on length for yellowfin tuna in the EPO from (Wild 1986). The size of the dots represents the number if fish (n) for each age (in quarters), by 10-cm intervals. The black line is the fixed growth assumption estimated externally, assumed in the fixed-growth models. The gray shaded region represents variation in length-at-age assuming a CV=7.5% in of length at age (mean ± 1.96 standard deviations)

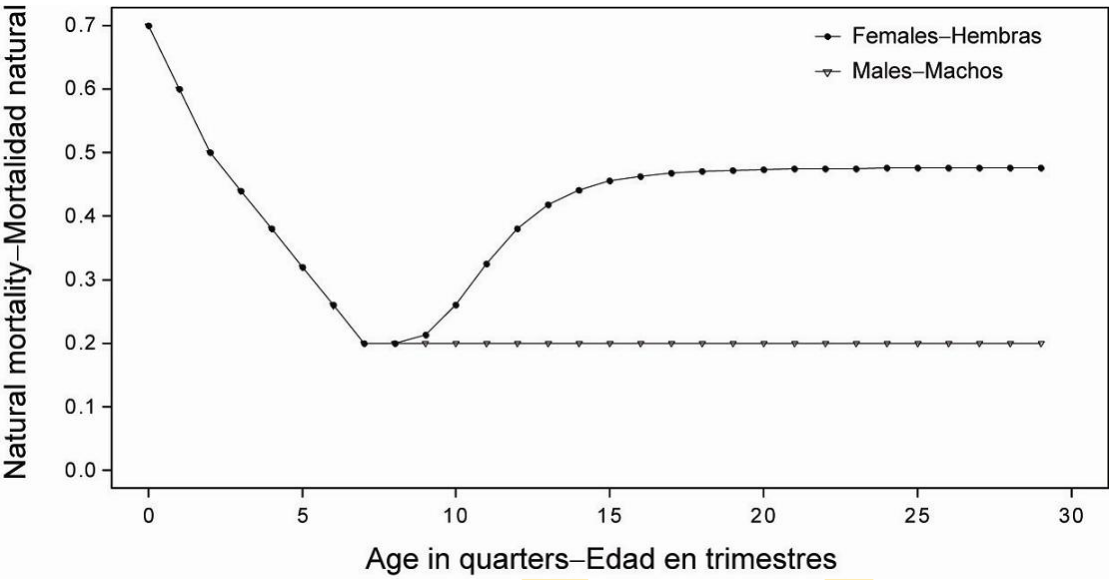

FIGURE 8. Natural mortality rates (*M*), by age and sex, at quarterly intervals, used for the assessment of yellowfin tuna in the EPO.

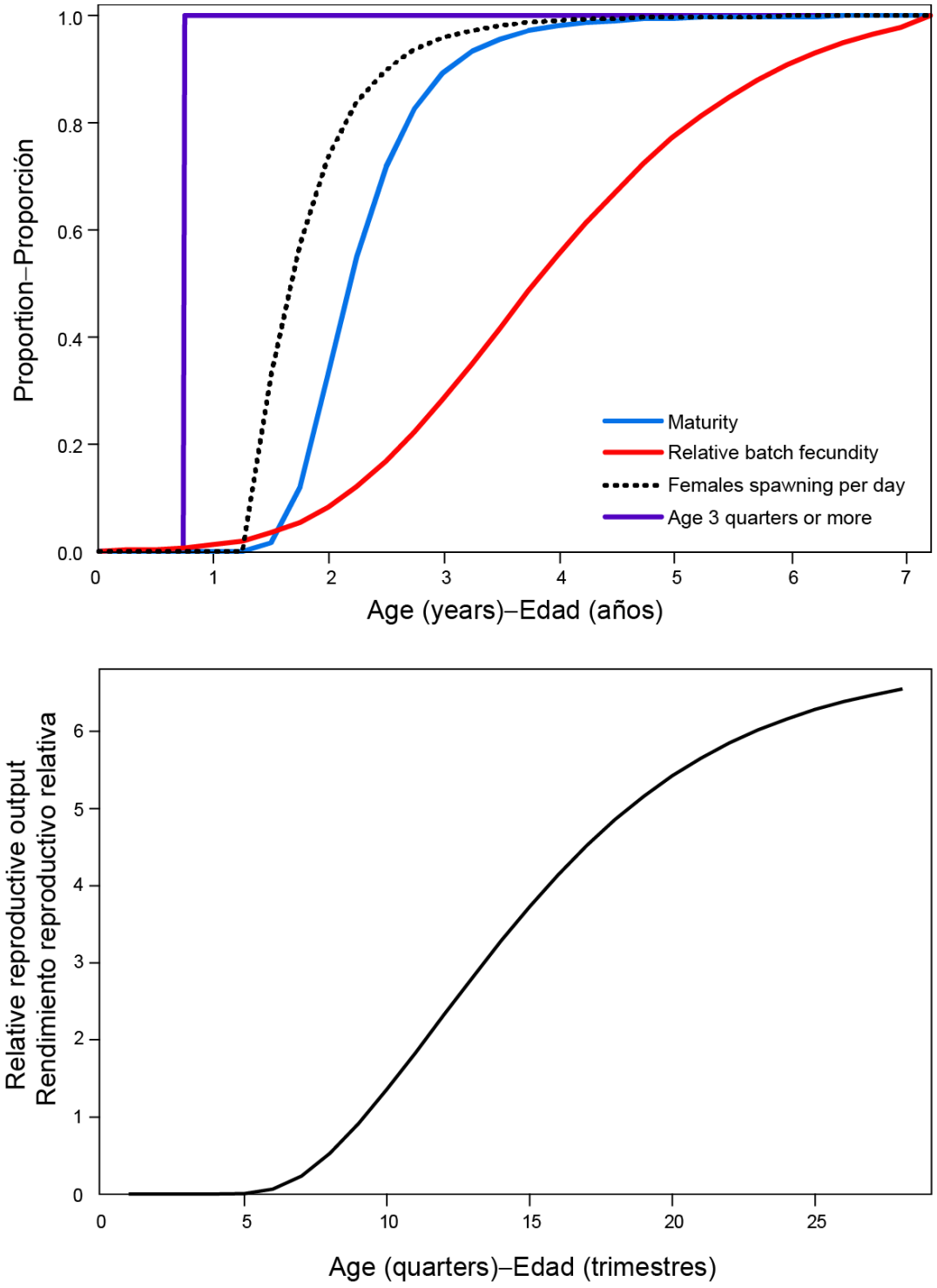

FIGURE 9. (**Top**) Relative contribution of each age to the reproductive output component (scaled to a maximum of one) for yellowfin tuna in the EPO (from Schaefer 1998). (**Bottom**) Relative fecundity -at-age curve used to estimate the index of spawning biomass of yellowfin tuna in the EPO.

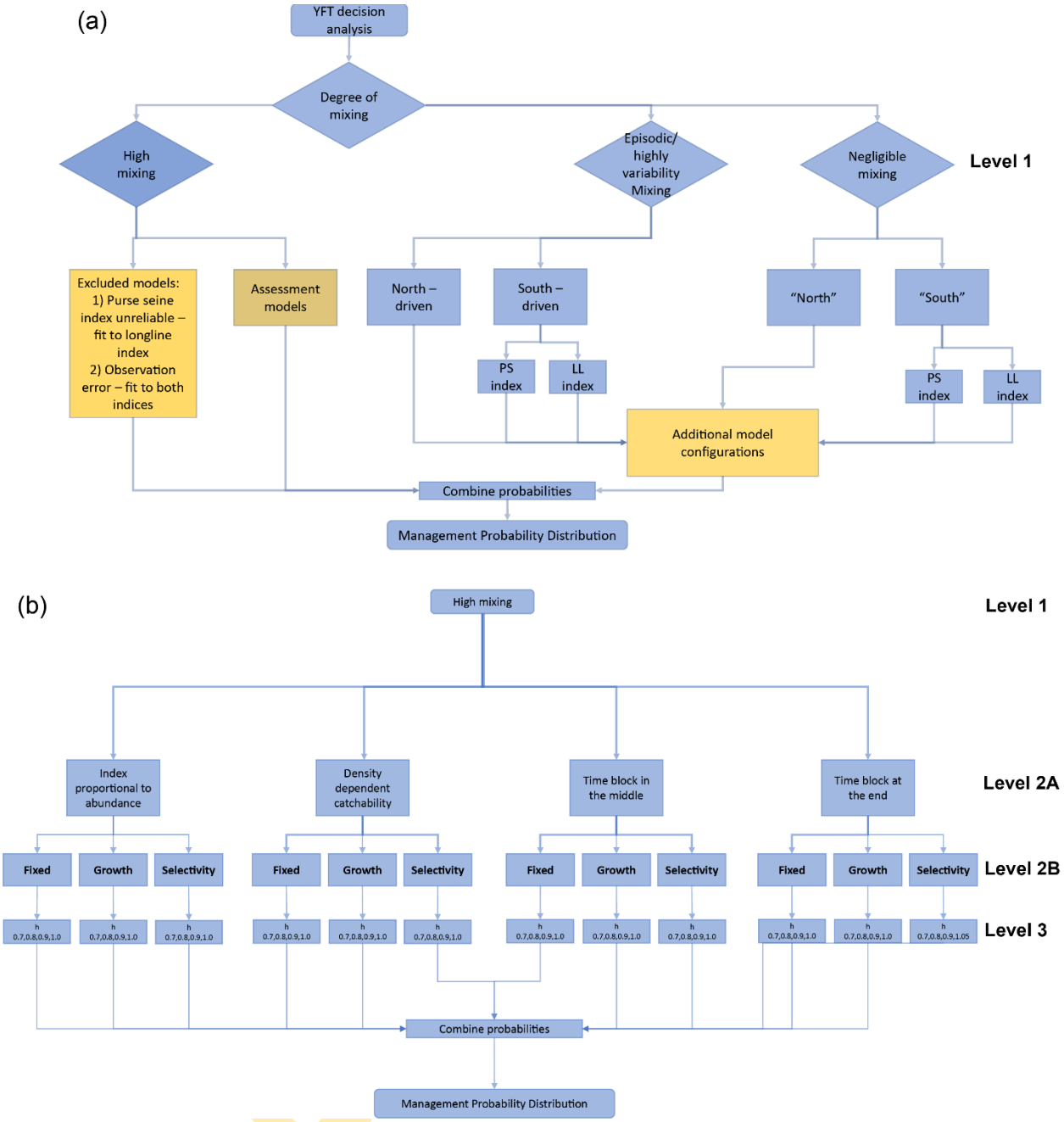

FIGURE 10. (**a**) Complete flow chart of hypotheses and models considered for the yellowfin risk assessment, and (**b**) flow chart of hypotheses and models included in the risk analysis (see text for details).

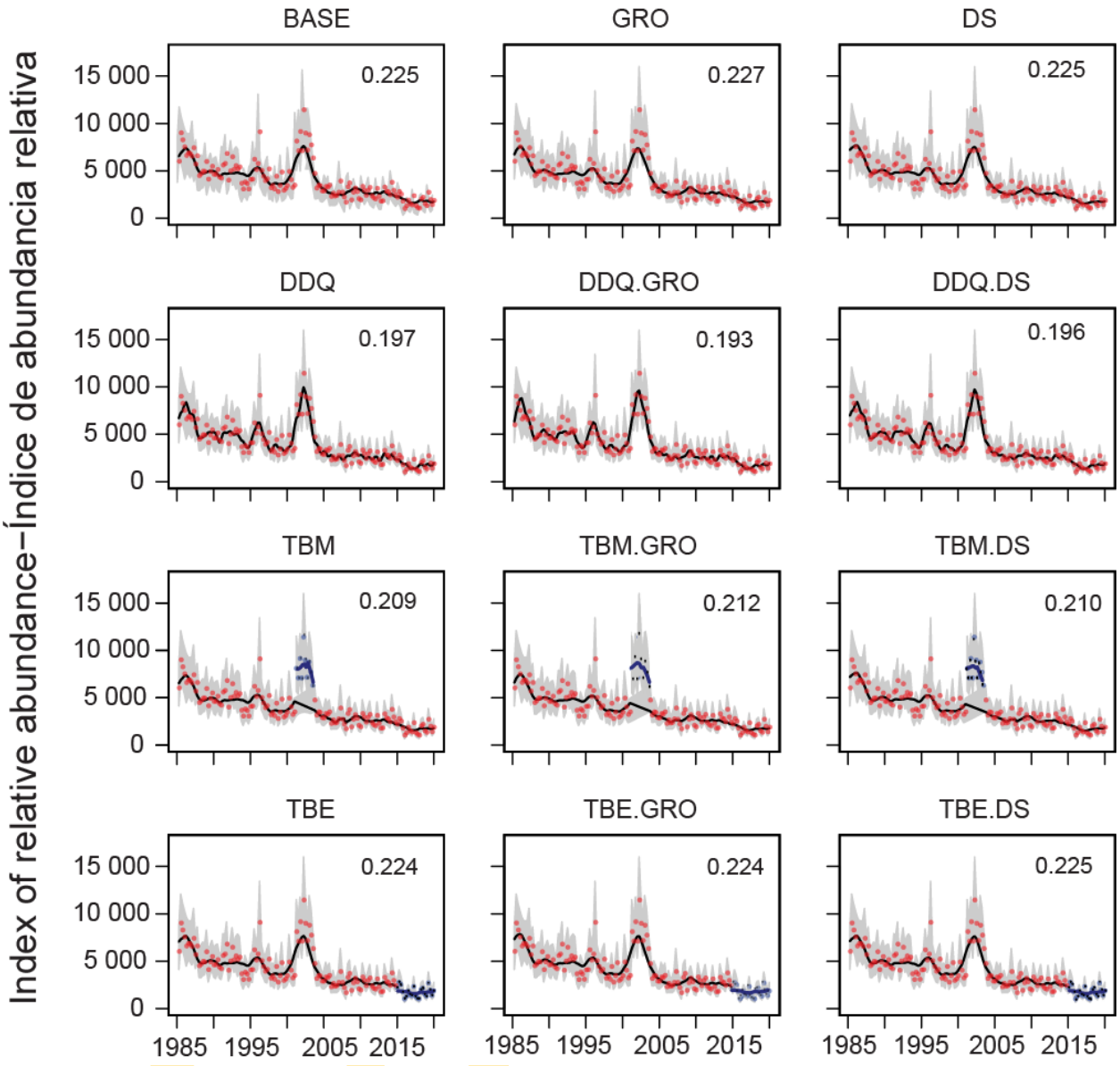

FIGURE 11. Model fits to the CPUE-based indices of abundance for the dolphin associated fisheries. The black lines are the estimated indices and the gray shading shows the approximate 95% confidence intervals (see text for details). The colored dots indicate the observed CPUE values, with blue dots indicating the data corresponding to time blocks (TBM and TBE) and the red dots the data outside of those time blocks. All models have steepness fixed at *h*=1. Model names are listed in Table 3.

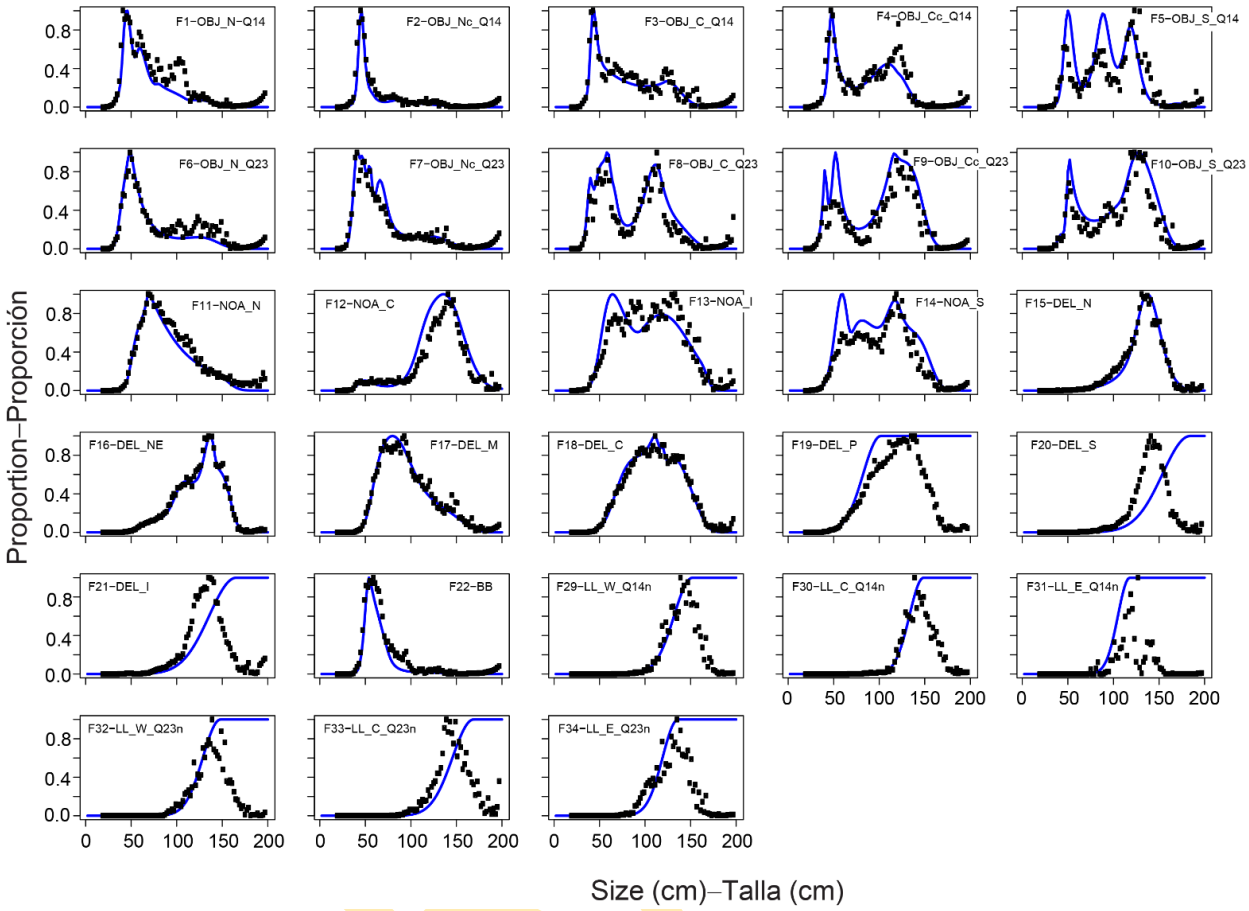

FIGURE 12. "Empirical" selectivities for the BASE *h*=1 model run (see text for details).

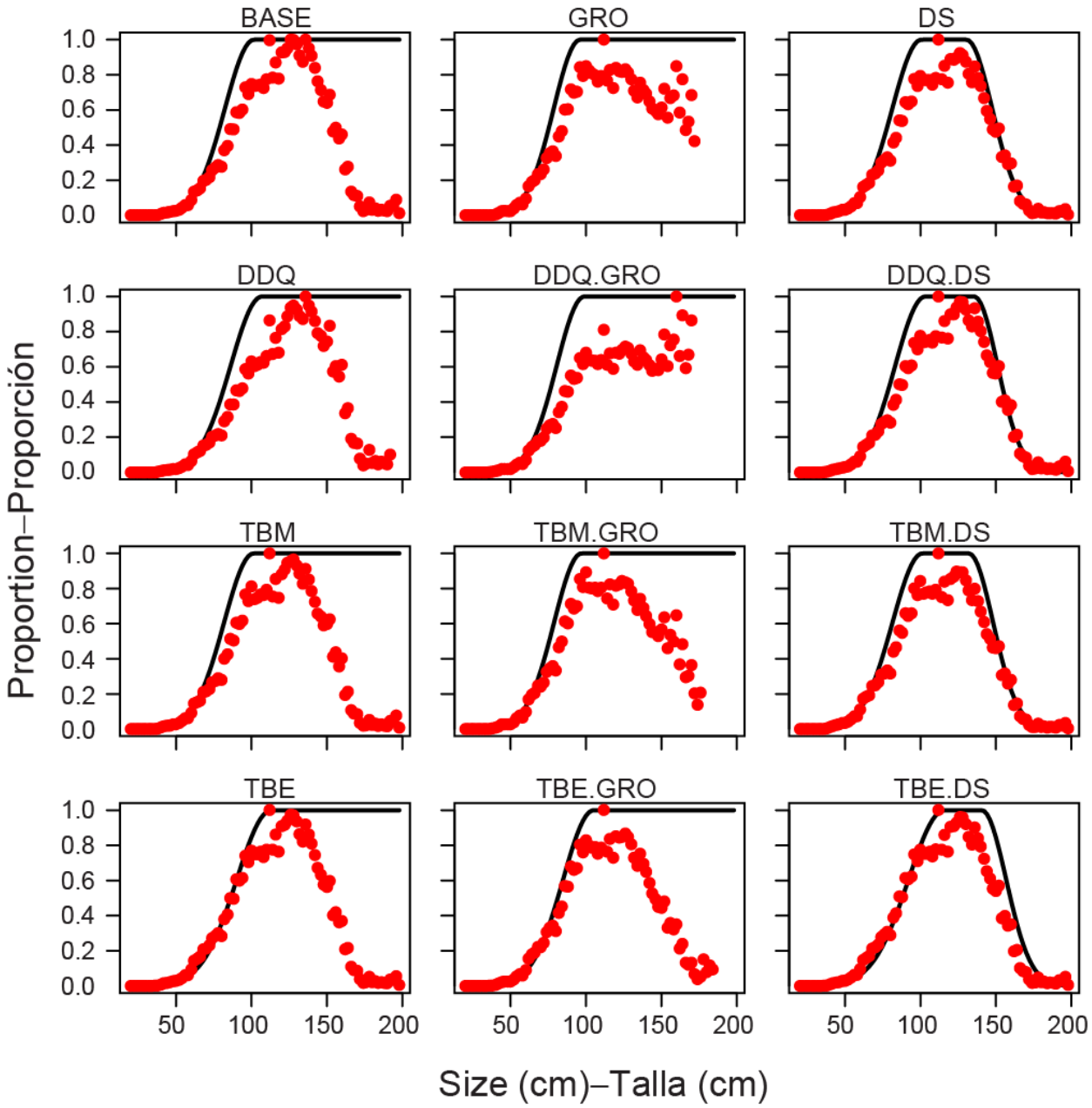

FIGURE 13. Comparison of estimated (black line) and empirical (red dots) average selectivity for fishery F-19-DEL_P in every reference model (with steepness h=1). The runs that do not have red dots in the larges size predicted zero fish in the population at that size, although there were fish in the sample.

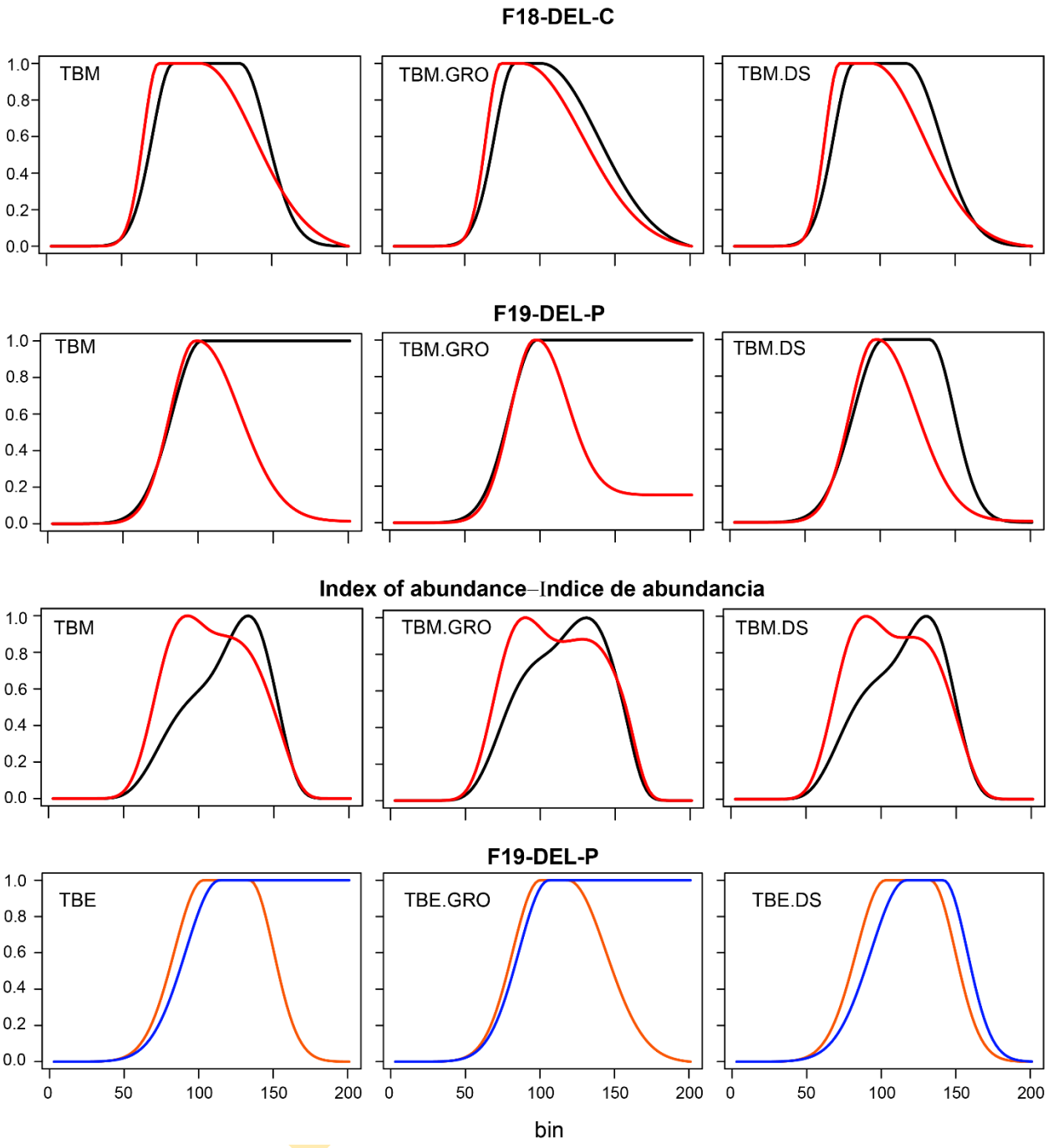

FIGURE 14. Estimated selectivity for the models with blocks in selectivity. The black line baseline selectivity, red line is the selectivity for 2002 quarter 3 to 2007 quarter 3. The blue line is the selectivity for 1984-2014, orange line is the selectivity for 2015-2019 (see model descriptions in Table 3)

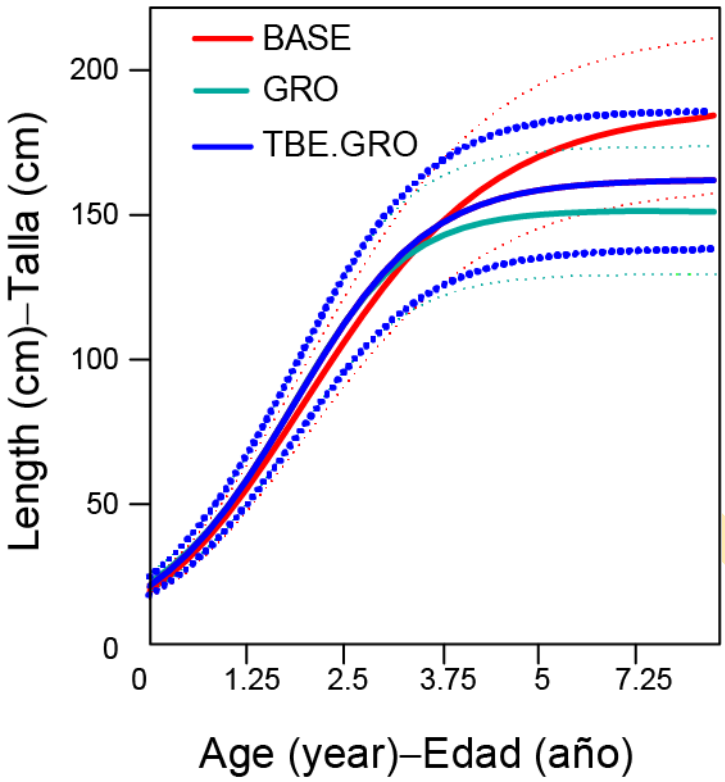

FIGURE 16. Schnute-Richard's growth curves estimated inside the integrated stock assessment models. The TBM.GRO curve is similar to the GRO curve and it is not shown. The solid lines are the mean length-at-age, the dotted lines represent variation in length-at-age, as the mean ± 1.96 standard deviations, which encompasses the size of 95% of the fish of a certain age in the population (assuming the length-at-age has a normal distribution)

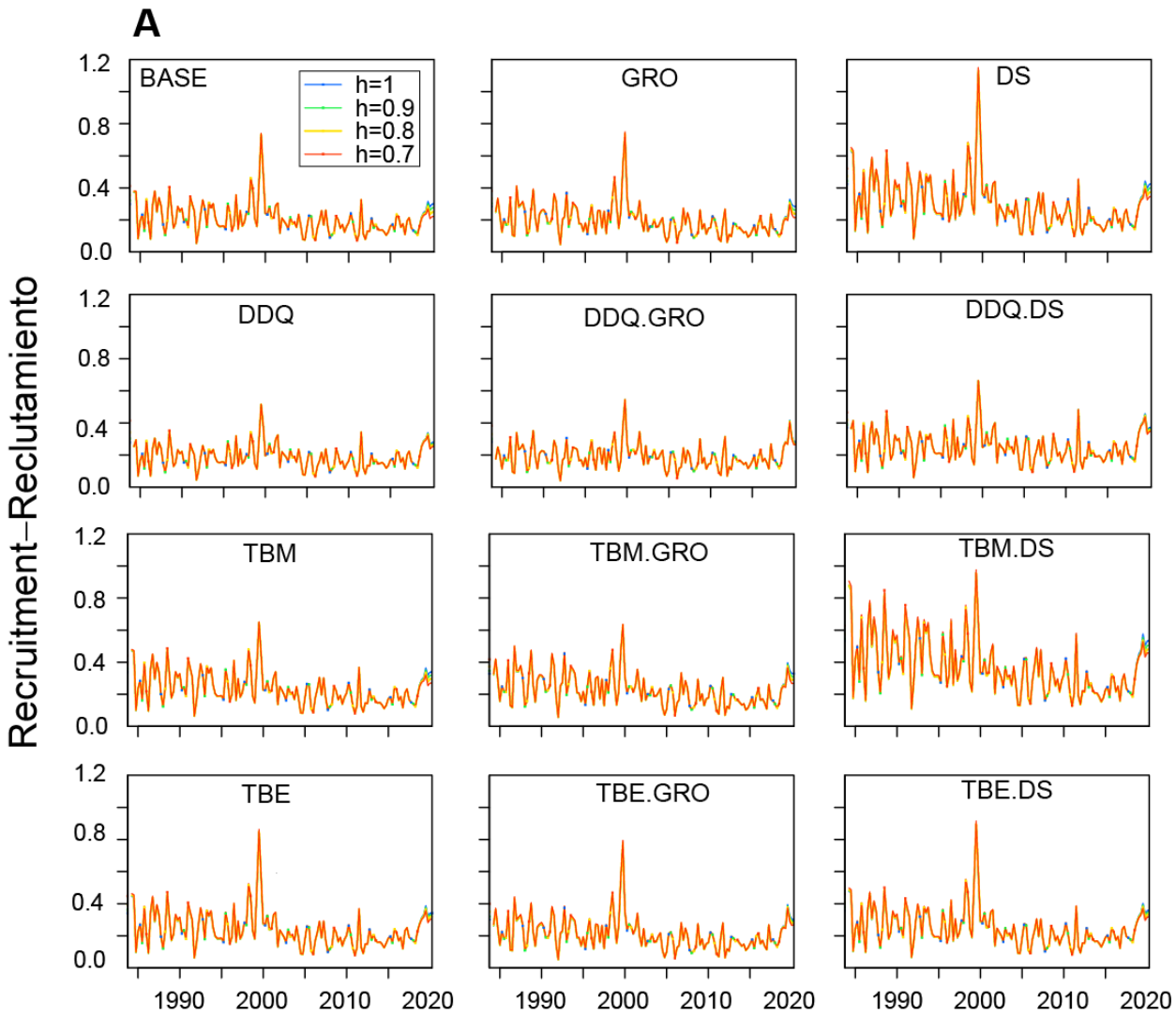

FIGURE 16. Estimated quarterly (**A.** Quarterly and **B (next page)** annual recruitment of yellowfin tuna to the fisheries of the EPO for the 48 models from the reference set. The lines indicate the maximum likelihood estimates (MLE) of recruitment (with colors corresponding to different values of the steepness parameter, *h*). In (B): the estimates are scaled so that the average recruitment is equal to 1.0 (dashed horizontal line), and the shaded areas indicate the approximate 95% confidence intervals around the estimates.

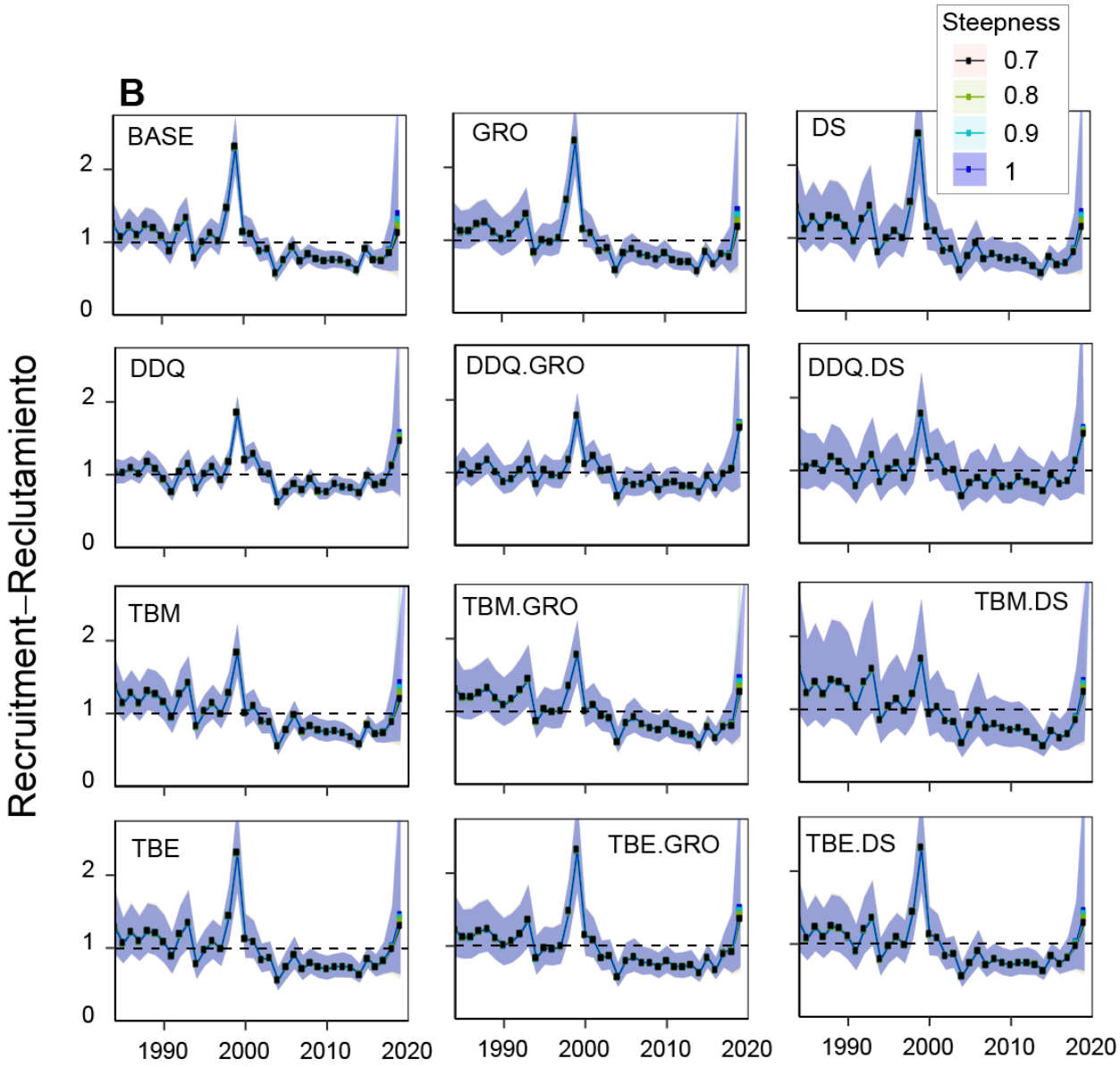

FIGURE 16 (cont).

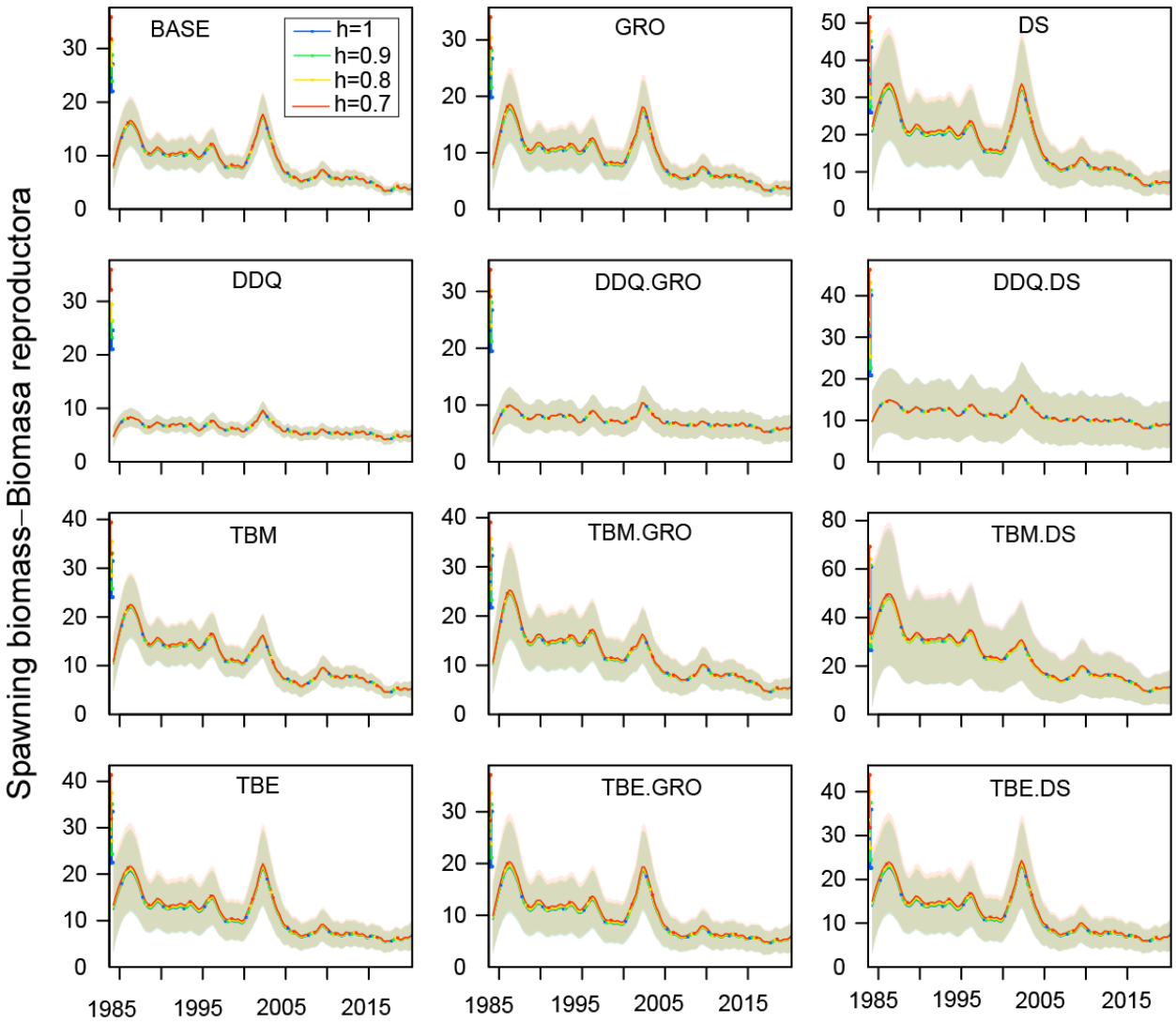

FIGURE 17. Index of spawning biomass (SBRs) for yellowfin tuna in the EPO for the 48 models from the reference set. The solid curves indicate the maximum likelihood estimates (with colors corresponding to different values of the steepness parameter, *h*). The shaded areas indicate the approximate 95% confidence intervals around those estimates. The colored bars and points at the beginning of each panel are the estimates of virgin spawning biomass for each model.

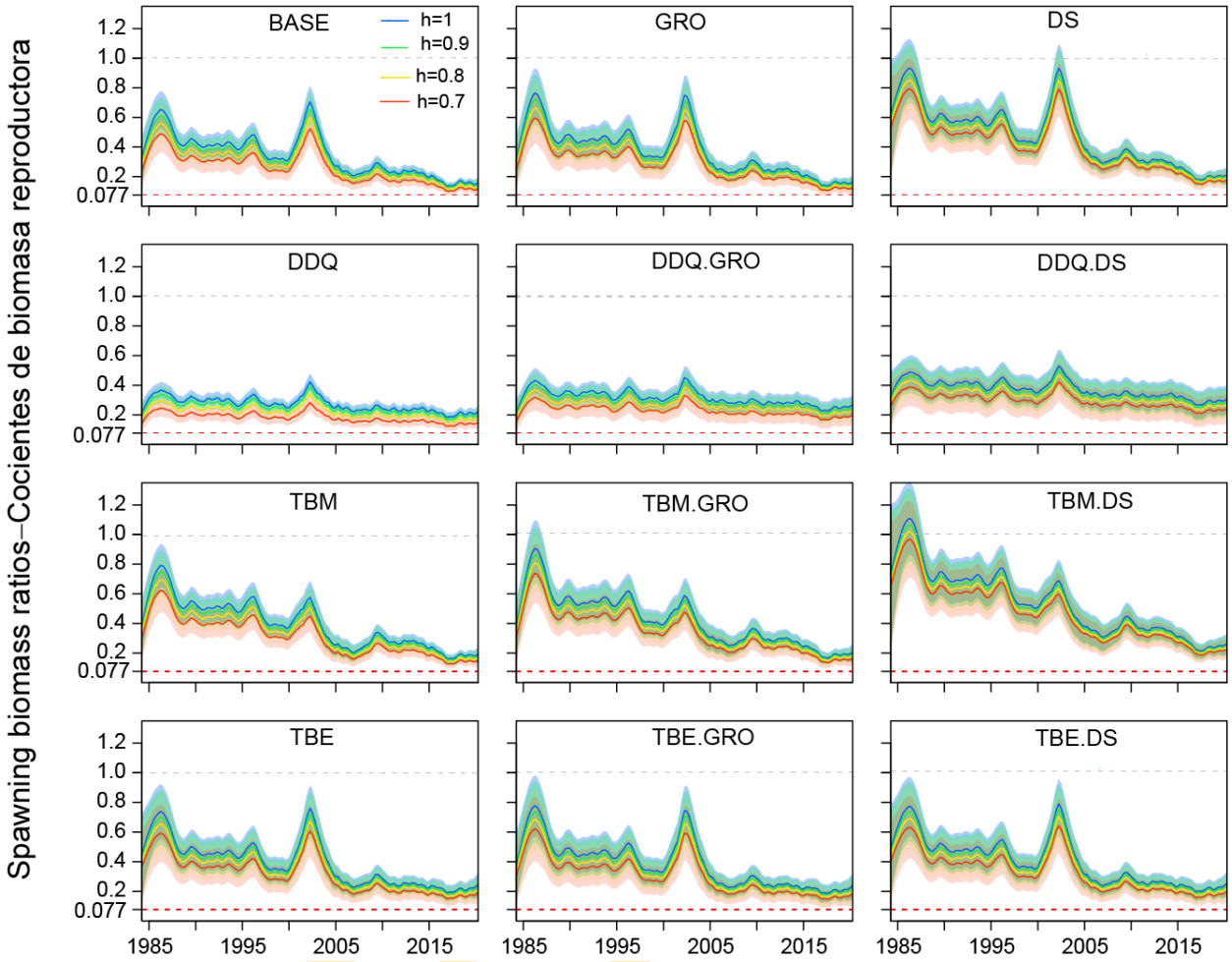

FIGURE 18. Spawning biomass ratios (SBRs) for yellowfin tuna in the EPO for the 48 models from the reference set. The red dashed horizontal line (at 0.077) identifies the SBR at the limit. The solid lines represent the maximum likelihood estimates (with colors corresponding to different values of the steepness parameter, *h*). The shaded areas are the approximate 95% confidence intervals around those estimates.

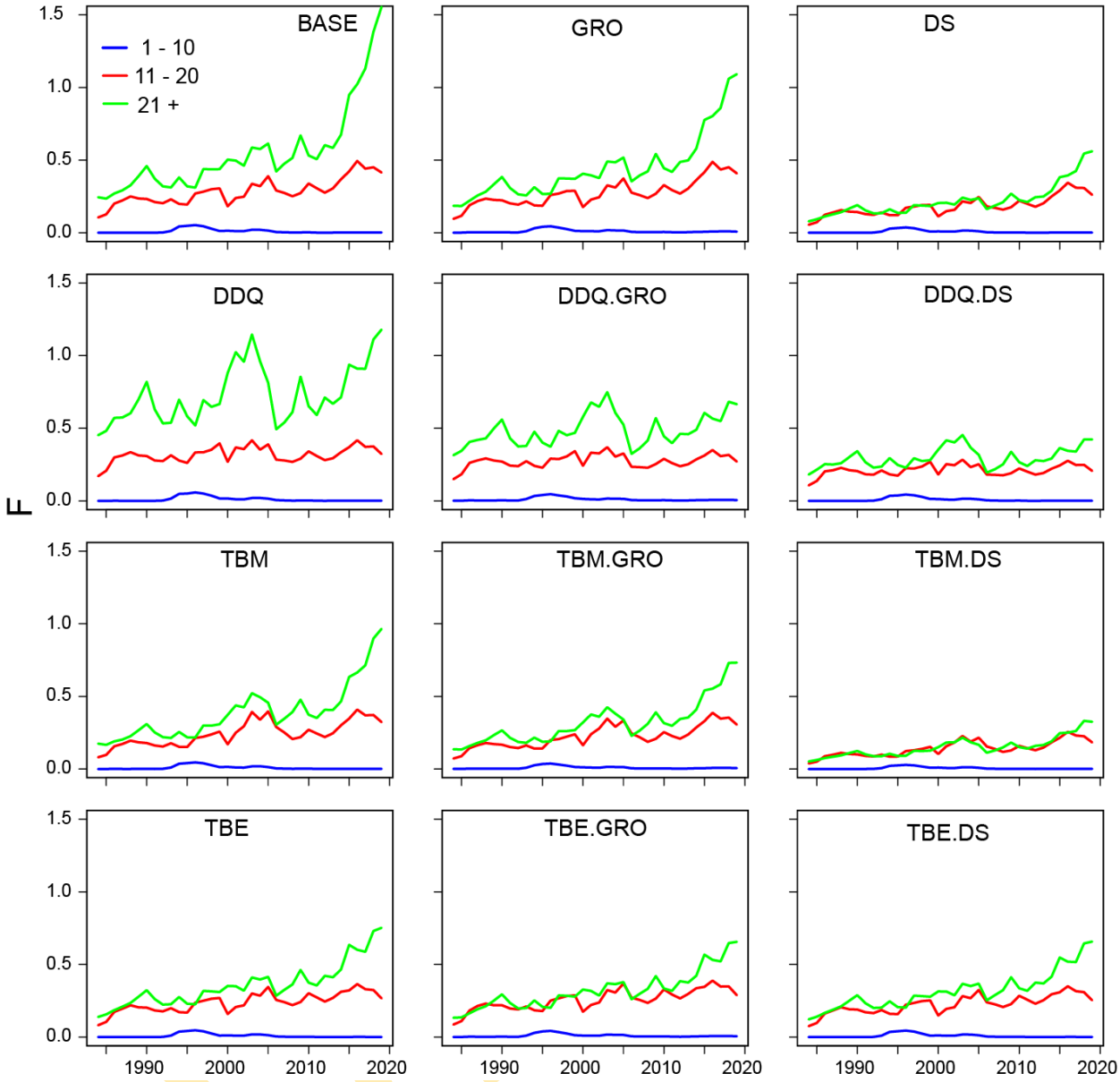

FIGURE 19. Average annual fishing mortality (*F*) by age groups, by all gears, of yellowfin tuna recruited to the fisheries of the EPO for all models with steepness *h*=1. The age groups are defined by age in quarters.

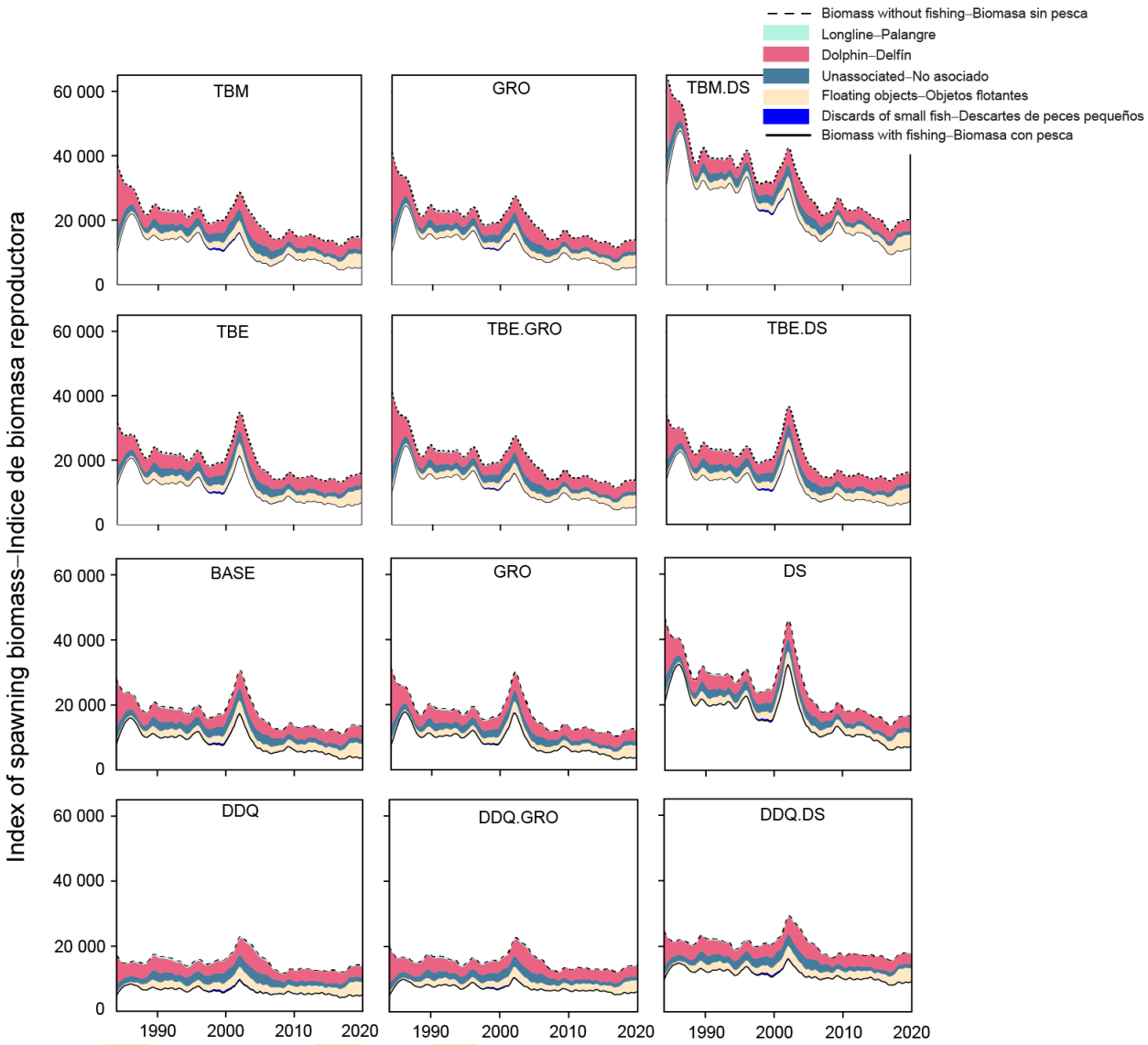

FIGURE 20. Fisheries impact plot: trajectory of the spawning biomass (a fecundity index, see text for details) of a simulated population of yellowfin tuna that was never exploited (top dashed line) and that predicted by the stock assessment model (bottom solid line). The shaded areas between the two lines show the portions of the impact attributed to each fishing method.

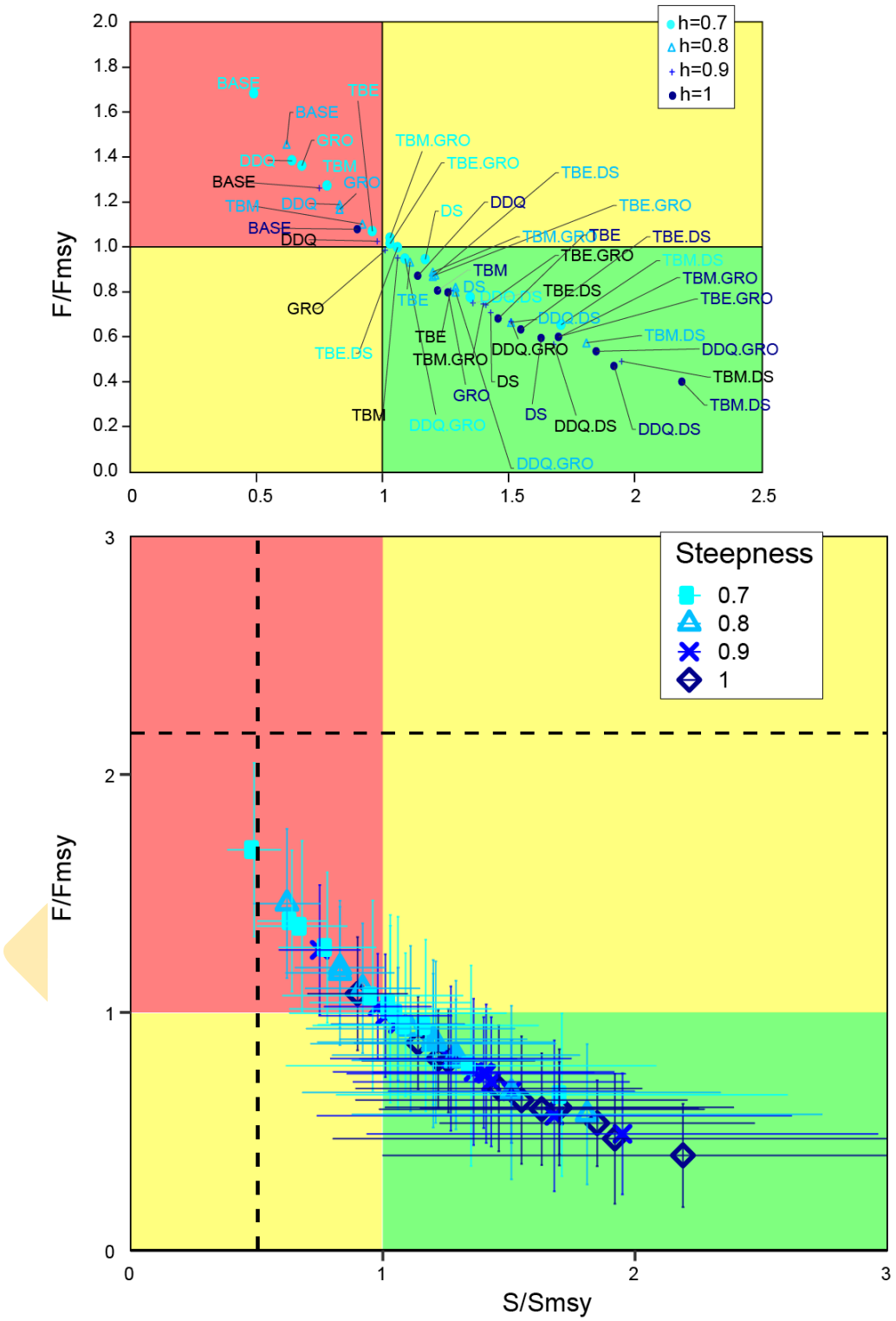

FIGURE 21. Kobe (phase) plot of the current spawning biomass and fishing mortality relative to their MSY reference points. The panels represent interim target reference points (S_{MSY_d} and F_{MSY}).(**Top**) Point estimate with the indication of the model. Each model is represented by a combination of its basic configuration plus the steepness assumption. (**Bottom**) 95 % confidence intervals around the estimates.

TABLE 1. Fisheries defined for the stock assessment of yellowfin tuna in the EPO. PS = purse seine; LP = pole and line; LL = longline; OBJ = sets on floating objects; NOA = sets on unassociated fish; DEL = sets on dolphin-associated schools. The areas that correspond to the fisheries are shown in Figure 1, and the discards are described in Section 2.2.3.

Fishery	Gear Type	Set type	Quarters	Catch data	Unit	Length Frequencies
F1-OBJ_N_Q14	PS	OBJ	1 and 4	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F2-OBJ_Nc_Q14	PS	OBJ	1 and 4	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F3-OBJ_C_Q14	PS	OBJ	1 and 4	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F4-OBJ_Cc_Q14	PS	OBJ	1 and 4	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F5-OBJ_S_Q14	PS	OBJ	1 and 4	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F6-OBJ_N_Q23	PS	OBJ	2 and 3	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F7-OBJ_Nc_Q23	PS	OBJ	2 and 3	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F8-OBJ_C_Q23	PS	OBJ	2 and 3	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F9-OBJ_Cc_Q23	PS	OBJ	2 and 3	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F10-OBJ_S_Q23	PS	OBJ	2 and 3	retained catch + discards from inefficiencies in fishing process	biomass	Fit
F11-NOA_N	PS	NOA	all	retained catch + discards	biomass	Fit
F12-NOA_C	PS	NOA	all	retained catch + discards	biomass	Fit
F13-NOA_I	PS	NOA	all	retained catch + discards	biomass	Fit
F14-NOA_S	PS	NOA	all	retained catch + discards	biomass	Fit
F15-DEL_N	PS	DEL	all	retained catch + discards	biomass	Fit
F16-DEL_NE	PS	DEL	all	retained catch + discards	biomass	Fit
F17-DEL_M	PS	DEL	all	retained catch + discards	biomass	Fit
F18-DEL_C	PS	DEL	all	retained catch + discards	biomass	Fit
F19-DEL_P	PS	DEL	all	retained catch + discards	biomass	Fit
F20-DEL_S	PS	DEL	all	retained catch + discards	biomass	Not Fit
F21-DEL_I	PS	DEL	all	retained catch + discards	biomass	Not Fit
F22-BB	LP		all	retained catch only	biomass	Fit
F25-OBJ_S_disc	PS	OBJ-Discard	all	discards of small fish from size-sorting the catch in the floating object fishery	biomass	
F26-OBJ_C_disc	PS	OBJ-Discard	all	discards of small fish from size-sorting the catch in the floating object fishery	biomass	

Fishery	Gear Type	Set type	Quarters	Catch data	Unit	Length Frequencies
F27-OBJ_I_disc	PS	OBJ-Discard	all	discards of small fish from size-sorting the catch in the floating object fishery	biomass	
F28-OBJ_N_disc	PS	OBJ-Discard	all	discards of small fish from size-sorting the catch in the floating object fishery	biomass	
F29-LL_W_Q14n	LL		1 and 4	retained catch only	numbers of fish (x 1000)	Not Fit
F30-LL_C_Q14n	LL		1 and 4	retained catch only	numbers of fish (x 1000)	Not Fit
F31-LL_E_Q14n	LL		1 and 4	retained catch only	numbers of fish (x 1000)	Not Fit
F32-LL_W_Q23n	LL		2 and 3	retained catch only	numbers of fish (x 1000)	Not Fit
F33-LL_C_Q23n	LL		2 and 3	retained catch only	numbers of fish (x 1000)	Not Fit
F34-LL_E_Q23n	LL		2 and 3	retained catch only	numbers of fish (x 1000)	Not Fit
F35-LL_W_Q14w	LL		1 and 4	retained catch only	biomass	
F36-LL_C_Q14w	LL		1 and 4	retained catch only	biomass	
F37-LL_E_Q14w	LL		1 and 4	retained catch only	biomass	
F38-LL_W_Q23w	LL		2 and 3	retained catch only	biomass	
F39-LL_C_Q23w	LL		2 and 3	retained catch only	biomass	
F40-LL_E_Q23w			2 and 3	retained catch only	biomass	
Indices	**Years**	**Gear**	**Quarters**	**Observation**	**Unit**	
S1-PS_DEL_VAST	1984-2019	PS		Used in the benchmark assessment	biomass	Fit
S2-LL_early_VAST_Q14	1984 - 1992	LL	1 and 4	Not used	numbers of fish (x 1000)	Not Fit
S2-LL_early_VAST_Q23	1984 - 1992	LL	2 and 3	Not used	numbers of fish (x 1000)	Not Fit
S2-LL_late_VAST_Q14	1995 - 2018	LL	1 and 4	Not used	numbers of fish (x 1000)	Not Fit
S2-LL_late_VAST_Q23	1995 - 2018	LL	2 and 3	Not used	numbers of fish (x 1000)	Not Fit

TABLE 2 Selectivity (at length unless noted otherwise) and weighting of composition data specified for the fisheries and surveys in the assessment. The asymptotic curves were modeled using a double-normal function. All selectivity at age were 1.0 for ages 1 to 29 (unless noted otherwise). The selectivity in grey were not estimated in the reference models.

Fishery	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
Selectivity	12-knot spline	11-knot spline	10-knot spline	10-knot spline	9-knot spline	9-knot spline
Data weighting	Francis/2	Francis	Francis/2	Francis/2	Francis/2	Francis
Fishery	**F7**	**F8**	**F9**	**F10**	**F11**	**F12**
Selectivity	6-knot spline	12-knot spline	14-knot spline	11-knot spline	10-knot spline	8-knot spline
Data weighting	Francis/2	Francis/2	Francis/2	Francis/2	Francis	Francis/2
Fishery	**F13**	**F14**	**F15**	**F16**	**F17**	**F18**
Selectivity	8-knot spline	9-knot spline	7-knot spline	10-knot spline	7-knot spline	Table 3
Data weighting	Francis/2	Francis/2	Francis	Francis/2	Francis	Francis
Fishery	**F19**	**F20**	**F21**	**F22**		
Selectivity	Double-normal	Asymptotic	Asymptotic	9-knot spline		
Data weighting	Francis	0	0	Francis		
Fishery	**F25**	**F26**	**F27**	**F28**	**F29**	**F30**
Selectivity	ages 3-5	ages 3-5	ages 3-5	ages 3-5	Asymptotic	Asymptotic
Data weighting	N/A	N/A	N/A	N/A	0	0
Fishery	**F31**	**F32**	**F33**	**F34**	**F35**	**F36**
Selectivity	Asymptotic	Asymptotic	Asymptotic	Asymptotic	mirror	mirror
Data weighting	0	0	0	0	0	
Fishery	**F37**	**F38**	**F39**	**F40**		
Selectivity	mirror	mirror	mirror	mirror		
Data weighting	0	0	0	0	0	
Survey	**S41**	**S23**	**S24**	**S42**	**S43**	
Selectivity	5-knot spline	Double-normal	Double-normal	Double-normal	Double-normal	
Data weighting	Francis	0	0	0	0	

TABLE 3. The reference set of models for the benchmark assessment of the yellowfin tuna in the EPO

	Model	Growth	Catchability of index	Index selectivity	Selectivity F19	Selectivity F18	Auxiliary data	Link
1.	BASE	Fixed	constant	constant	Asymptotic	Dome-shape (11-knot spline)		html
2.	GRO	Estimated	constant	constant	Asymptotic		age-at-length	html
3.	DS	Fixed	constant	constant	Dome-shape			html
4.	DDQ	Fixed	Density-dependent	constant	Asymptotic			html
5.	DDQ.GRO	Estimated	Density-dependent	constant	Asymptotic		age-at-length	html
6.	DDQ.DS	Fixed	Density-dependent	constant	Dome-shape			html
7.	TBM	Fixed	Blocks: baseline, 2001-2003.Q2	block: 1984 – 2002.Q2 2002.Q3-2007.Q3	Blocks: Asymptotic (baseline) Dome-shape (2002.Q3-2007.Q3)	Double-normal Block: Dome-shape (baseline) Dome-shape (2002.Q3-2007.Q3)		html
8.	TBM.GRO	Estimated	Blocks: baseline, 2001-2003.Q2	block: 1984 – 2002.Q2 2002.Q3-2007.Q3			age-at-length	html
9.	TBM.DS	Fixed	Blocks: baseline, 2001-2003.Q2	block: 1984 – 2002.Q2 2002.Q3-2007.Q3				html
10.	TBE	Fixed	Blocks: baseline, 2001-2003.Q2	constant	Blocks: Dome-shape (1984-2014), Asymptotic (2015-2019)	Dome-shape (11-knot spline)		html
11.	TBE.GRO	Estimated	Blocks: baseline, 2001-2003.Q2	constant			age-at-length	html
12.	TBE.DS	Fixed	Blocks: baseline, 2001-2003.Q2	constant				html

TABLE 4. The maximum gradient of the all model runs. NA: run does not converge/Hessian is not positive definite.

	1	2	3	4	5	6	7	8	9	10	11	12
Steepness	**BASE**	**GRO**	**DS**	**DDQ**	**DDQ.GRO**	**DDQ.DS**	**TBM**	**TBM.GRO**	**TBM.DS**	**TBE**	**TBE.GRO**	**TBE.DS**
h = 1	1.28E-03	1.40E-04	5.38E-05	2.41E-04	3.48E-05	1.63E-04	2.24E-03	3.45E+00	1.05E+01	2.21E-04	8.28E-05	1.42E-04
h = 0.9	7.36E-05	7.30E-05	6.29E-05	5.52E-04	1.09E-02	3.02E-03	5.12E-04	3.53E+00	1.10E+00	2.77E-05	4.13E-03	9.12E-05
h = 0.8	9.35E-05	1.32E-02	1.37E-04	2.39E-04	2.51E-05	1.62E-03	1.58E-04	1.12E+01	1.27E+00	4.51E-04	2.86E-05	3.96E-04
h = 0.7	2.48E-04	4.12E-04	6.67E-05	2.85E-02	1.30E-03	6.18E-04	1.02E-03	4.41E+00	9.80E+00	1.61E-03	8.23E-04	1.41E-03

TABLE 5 Root mean square error (RMSE) and negative log-likelihood (NLL) for the index of abundance for all model runs.

	1	2	3	4	5	6	7	8	9	10	11	12
Steepness	**BASE**	**GRO**	**DS**	**DDQ**	**DDQ.GRO**	**DDQ.DS**	**TBM**	**TBM.GRO**	**TBM.DS**	**TBE**	**TBE.GRO**	**TBE.DS**
RMSE												
h = 1	0.23	0.23			0.19	0.20	0.21	0.21	0.21	0.22	0.22	0.22
h = 0.9	0.23	0.23	0.22	0.22	0.20	0.20	0.21	0.21	0.21	0.22	0.22	0.22
h = 0.8	0.23	0.23	0.23	0.20	0.19	0.20	0.21	0.21	0.21	0.22	0.22	0.22
h = 0.7	0.23	0.23	0.23	0.20	0.19	0.20	0.21	0.21	0.21	0.22	0.22	0.22
NLL differences to the lowest (NLL=-164.85)												
h = 1	32.3	34.8	32.3	4.5	0.0	2.5	16.0	17.3	16.2	31.8	32.1	32.0
h = 0.9	32.3	34.8	32.3	4.5	0.1	2.5	16.0	17.3	16.2	31.8	32.1	32.0
h = 0.8	32.4	34.9	32.4	4.5	0.1	2.6	16.0	17.4	16.3	31.9	32.2	32.0
h = 0.7	32.5	35.2	32.5	4.5	0.1	2.6	16.0	17.4	16.3	32.0	32.4	32.2

TABLE 6. Average adjusted input sample size (n adj) and average effective sample size for each fishery and for the survey, by model configuration with steepness h=1. LP = pole and line; LL = longline; OBJ = sets on floating objects; NOA = sets on unassociated fish; DEL = sets on dolphin-associated schools. The fisheries area defined in Table 2.1 and areas are shown in Figure 2.1. The cells in gray are of length composition data not use in the model. The names in bold are the fishery with the largest ranges in effective sample size.

		Effective n												
	n adj	**BASE**	**DDQ**	**DDQ.DS**	**DDQ.GRO**	**DS**	**GRO**	**TBE**	**TBE.DS**	**TBE.GRO**	**TBM**	**TBM.DS**	**TBM.GRO**	**Range**
Survey	11.8	59	62	63	63	61	60	**64**	**64**	**65**	**64**	**64**	**63**	**7**
Fishery														
F1-OBJ_N-Q14	7.3	63	64	63	61	62	61	62	62	61	63	63	61	3
F2-OBJ_Nc_Q14	6.1	36	35	35	35	36	35	36	36	35	36	36	35	1
F3-OBJ_C_Q14	9.4	58	59	59	72	58	70	59	59	70	59	59	72	**14**
F4-OBJ_Cc_Q14	4.0	33	33	33	32	33	32	33	33	32	33	33	32	2
F5-OBJ_S_Q14	3.6	40	39	38	38	40	40	41	41	40	40	39	39	3
F6-OBJ_N_Q23	9.4	83	83	82	87	83	88	83	83	88	83	83	87	6
F7-OBJ_Nc_Q23	4.1	38	37	37	35	38	36	39	39	36	38	38	35	4
F8-OBJ_C_Q23	6.8	55	55	54	52	55	53	55	55	52	55	55	53	3
F9-OBJ_Cc_Q23	3.8	30	30	30	31	30	32	30	30	31	30	30	31	2
F10-OBJ_S_Q23	4.8	25	26	26	25	25	24	25	25	25	26	26	25	1
F11-NOA_N	5.3	63	61	61	61	63	62	63	63	62	63	62	62	2
F12-NOA_C	2.3	25	25	25	24	25	25	25	25	25	25	25	25	1
F13-NOA_I	2.0	33	33	33	34	34	34	34	34	34	34	34	34	1
F14-NOA_S	3.0	22	21	22	21	22	21	22	22	21	22	22	21	1
F15-DEL_N	4.1	108	107	106	105	109	107	109	109	109	109	109	108	4
F16-DEL_NE	7.5	113	113	113	114	114	115	114	114	116	114	115	115	3
F17-DEL_M	3.8	53	53	53	53	53	54	53	53	54	53	53	53	1
F18-DEL_C	7.8	159	161	162	162	160	160	160	160	160	155	157	155	**8**
F19-DEL_P	6.2	105	110	111	111	110	107	110	110	111	108	114	111	**9**
F20-DEL_S	*4.2*	61	69	48	101	44	103	52	50	66	53	40	92	63
F21-DEL_I	*4.5*	45	47	39	50	37	50	40	39	41	41	35	47	15
F22-BB	*5.2*	45	45	45	50	45	50	45	45	50	45	45	49	5
F29-LL_W_Q14n	*6.9*	44	49	41	55	38	54	42	41	45	42	37	52	18

		Effective n												
	n adj	**BASE**	**DDQ**	**DDQ.DS**	**DDQ.GRO**	**DS**	**GRO**	**TBE**	**TBE.DS**	**TBE.GRO**	**TBM**	**TBM.DS**	**TBM.GRO**	**Range**
F30-LL_C_Q14n	*8.9*	58	67	55	77	49	76	54	53	62	55	47	74	31
F31-LL_E_Q14n	*2.8*	18	19	17	18	16	18	17	17	17	18	17	18	2
F32-LL_W_Q23n	*5.0*	47	52	45	52	42	49	45	44	45	45	40	48	12
F33-LL_C_Q23n	*3.4*	31	35	28	44	26	41	28	28	31	28	25	38	19
F34-LL_E_Q23n	*2.0*	23	26	23	24	21	23	22	22	22	23	21	23	4

TABLE 7 Number of estimated parameters, negative log-likelihood (NLL): (1) without the age at length data and (2) only the age at length data

Number of estimated parameters												
Steepness	**BASE**	**GRO**	**DS**	**DDQ**	**DDQ.GRO**	**DDQ.DS**	**TBM**	**TBM.GRO**	**TBM.DS**	**TBE**	**TBE.GRO**	**TBE.DS**
h = 1	332	336	334	333	337	335	339	343	341	341	345	343
h = 0.9	332	336	334	333	337	335	339	343	341	341	345	343
h = 0.8	332	336	334	333	337	335	339	343	341	341	345	343
h = 0.7	332	336	334	333	337	335	339	343	341	341	345	343
(1) NLL without age at length data												
h = 1	2134.1	2112.3	2127.0	2107.3	2071.6	2079.7	2087.4	2069.8	2058.9	2127.0	2035.8	2125.9
h = 0.9	2133.7	2111.8	2126.6	2106.7	2071.6	2079.6	2086.8	2069.2	2058.3	2126.6	2035.5	2125.5
h = 0.8	2133.4	2111.5	2126.2	2106.1	2071.6	2079.6	2086.3	2068.7	2057.6	2126.2	2035.2	2125.1
h = 0.7	2133.4	2111.5	2126.1	2105.6	2071.6	2079.6	2086.0	2068.4	2056.8	2126.1	2035.1	2124.9
(2) NLL only age at length data												
h = 1	N/A	56.0	N/A	N/A	55.8	N/A	N/A	50.3	N/A	N/A	N/A	55.4
h = 0.9	N/A	56.0	N/A	N/A	55.8	N/A	N/A	50.3	N/A	N/A	N/A	55.4
h = 0.8	N/A	56.0	N/A	N/A	55.8	N/A	N/A	50.2	N/A	N/A	N/A	55.3
h = 0.7	N/A	55.8	N/A	N/A	55.8	N/A	N/A	50.2	N/A	N/A	N/A	55.2
Δ AIC – no age at length data				Min AIC=	4795.7							
h = 1	136.6	100.9	87.0	55.3	21.5	33.7	57.1	30.0	4.1	140.2	131.7	142.0
h = 0.9	135.7	99.8	85.7	55.2	21.5	33.6	56.0	28.8	2.8	139.4	130.9	141.2
h = 0.8	135.1	99.2	84.5	55.1	21.5	33.5	54.9	27.7	1.4	138.8	130.4	140.5
h = 0.7	135.2	99.4	83.5	55.2	21.6	33.5	54.3	27.0	0.0	138.5	130.3	140.1

TABLE 8. Ratio between catchabilities of the index of abundance in the block and the baseline and non-linearity coefficient *c*

Steepness	**$q_{2015\text{-}2019}/q$**		
	TBE	TBE.GRO	TBE.DS
h = 1	0.91	0.92	0.86
h = 0.9	0.92	0.92	0.86
h = 0.8	0.92	0.93	0.87
h = 0.7	0.92	0.93	0.87
	$q_{2001\text{-}2003.Q2}/q$		
	TBM	TBM.GRO	TBM.DS
h = 1	1.64	1.82	1.74
h = 0.9	1.64	1.81	1.74
h = 0.8	1.65	1.81	1.74
h = 0.7	1.65	1.82	1.75
		c	
	DDQ	DDQ.GRO	DDQ.DS
h = 1	1.7	2.2	2.1
h = 0.9	1.7	2.2	2.1
h = 0.8	1.7	2.2	2.0
h = 0.7	1.7	2.2	2.0

TABLE 9. The management table for yellowfin in the Eastern Pacific Ocean. $S_{current}$ and S_{MSY_d} are the spawning biomass at the beginning of 2020 and at dynamic MSY level, respectively. $F_{current}$ and F_{MSY} are the fishing mortality during 2017-2019 and at MSY, respectively. S_{LIMIT} and F_{LIMIT} are the limit reference points for spawning biomass and fishing morality, respectively. $C_{current}$ is the total catch (in metric tons) of yellowfin in 2019 and MSY_d is the dynamic MSY.

	1	2	3	4	5	6	7	8	9	10	11	12
	BASE	**GRO**	**DS**	**DDQ**	**DDQ.GRO**	**DDQ.DS**	**TBM**	**TBM.GRO**	**TBM.DS**	**TBE**	**TBE.GRO**	**TBE.DS**
***h* = 1.0**												
MSY	461,752	488,404	586,672	425,788	466,324	511,876	497,760	543,960	710,188	494,796	509,932	510,824
MSY_d	257,732	263,175	290,662	271,054	299,762	319,271	269,331	288,203	353,699	290,869	300,961	297,008
$C_{current}/MSY\_d$	0.97	0.95	0.87	0.92	0.83	0.79	0.93	0.87	0.72	0.86	0.83	0.85
S_{MSY}/S_0	0.32	0.24	0.27	0.31	0.23	0.27	0.30	0.23	0.26	0.29	0.24	0.29
$S_{current}/S_0$	0.15	0.16	0.21	0.22	0.27	0.30	0.19	0.20	0.26	0.24	0.24	0.25
$S_{current}/S_{LIMIT}$	2.00	2.09	2.71	2.84	3.45	3.93	2.47	2.62	3.37	3.17	3.05	3.26
$p(S_{current}<S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.40	0.40	0.27	0.33	0.28	0.22	0.33	0.31	0.20	0.28	0.30	0.27
$p(F_{current}> F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	0.90	1.26	1.63	1.14	1.85	1.92	1.22	1.70	2.19	1.46	1.70	1.55
$p(S_{current}<S_{MSY\_d})$	0.84	0.07	0.03	0.14	0.00	0.05	0.08	0.01	0.03	0.06	0.02	0.05
$F_{current}/F_{MSY}$	1.08	0.80	0.59	0.87	0.53	0.47	0.81	0.60	0.40	0.68	0.60	0.63
$p(F_{recent}> F_{MSY})$	0.74	0.03	0.00	0.10	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00
***h* = 0.9**												
MSY	468,040	481,752	573,148	436,744	459,168	501,548	496,352	528,252	677,592	493,256	501,144	506,556
MSY_d	260,403	252,946	267,120	267,881	276,496	293,116	259,476	263,425	308,512	276,548	278,752	279,319
$C_{current}/MSY\_d$	0.96	0.98	0.94	0.93	0.90	0.86	0.97	0.95	0.82	0.91	0.90	0.90
S_{MSY}/S_0	0.35	0.29	0.31	0.35	0.28	0.31	0.33	0.28	0.30	0.33	0.29	0.32
$S_{current}/S_0$	0.14	0.15	0.20	0.20	0.25	0.29	0.18	0.19	0.25	0.23	0.22	0.24
$S_{current}/S_{LIMIT}$	1.86	1.95	2.59	2.61	3.23	3.73	2.32	2.47	3.24	2.97	3.05	3.06
$p(S_{current}<S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.47	0.46	0.31	0.38	0.32	0.25	0.38	0.36	0.23	0.33	0.30	0.31
$p(F_{current}> F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	0.75	1.01	1.43	0.98	1.51	1.68	1.06	1.40	1.95	1.27	1.41	1.36
$P(S_{current}<S_{MSY\_d})$	1.00	0.47	0.06	0.57	0.02	0.08	0.33	0.04	0.03	0.13	0.07	0.10
$F_{current}/F_{MSY}$	1.26	0.99	0.71	1.02	0.67	0.57	0.95	0.75	0.49	0.81	0.74	0.75
$p(F_{current}> F_{MSY})$	0.97	0.46	0.02	0.59	0.00	0.00	0.34	0.02	0.00	0.10	0.04	0.06
***h* = 0.8**												
MSY	483,904	485,012	565,840	462,136	463,640	498,952	502,580	521,748	658,140	499,520	502,460	509,704

	1	2	3	4	5	6	7	8	9	10	11	12
	BASE	**GRO**	**DS**	**DDQ**	**DDQ.GRO**	**DDQ.DS**	**TBM**	**TBM.GRO**	**TBM.DS**	**TBE**	**TBE.GRO**	**TBE.DS**
MSY_d	269,568	251,063	249,703	271,954	261,577	272,308	254,710	248,137	276,016	268,398	266,591	267,751
$C_{current}$/MSY_d	0.92	0.99	1.01	0.92	0.95	0.93	0.98	1.01	0.92	0.94	0.94	0.94
S_{MSY}/S_0	0.37	0.32	0.34	0.37	0.32	0.34	0.36	0.32	0.33	0.36	0.32	0.35
$S_{current}/S_0$	0.13	0.14	0.19	0.18	0.23	0.27	0.16	0.18	0.24	0.21	0.20	0.22
$S_{current}/S_{LIMIT}$	1.68	1.79	2.44	2.31	2.93	3.46	2.13	2.30	3.07	2.74	2.63	2.83
$p(S_{current}<S_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.55	0.53	0.36	0.45	0.37	0.29	0.44	0.42	0.26	0.38	0.41	0.37
$p(F_{currentt}> F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	0.62	0.83	1.29	0.83	1.29	1.51	0.92	1.20	1.81	1.11	1.21	1.20
$p(S_{current}<S_{MSY\_d})$	1.00	0.94	0.12	0.97	0.08	0.11	0.75	0.15	0.04	0.30	0.19	0.21
$F_{current}/F_{MSY}$	1.46	1.17	0.82	1.19	0.80	0.66	1.10	0.89	0.57	0.93	0.88	0.87
$p(F_{current}> F_{MSY})$	1.00	0.86	0.13	0.93	0.06	0.04	0.77	0.21	0.00	0.35	0.24	0.23
***h* = 0.7**												
MSY	518,192	502,584	566,512	521,896	488,020	508,960	521,792	526,380	650,584	518,396	517,428	524,164
MSY_d	289,293	256,702	235,527	291,255	254,438	255,332	255,934	238,816	248,957	266,352	262,019	261,308
$C_{current}$/MSY_d	0.86	0.97	1.07	0.86	0.98	0.99	0.98	1.04	1.02	0.94	0.96	0.96
S_{MSY}/S_0	0.40	0.35	0.37	0.40	0.35	0.37	0.38	0.35	0.36	0.38	0.35	0.38
$S_{current}/S_0$	0.11	0.12	0.17	0.15	0.19	0.24	0.15	0.16	0.22	0.19	0.18	0.20
$S_{current}/S_{LIMIT}$	1.45	1.58	2.26	1.90	2.53	3.10	1.90	3.21	2.89	2.44	2.35	2.55
$p(S_{current}<S_{LIMIT})$	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{current}/F_{LIMIT}$	0.65	0.62	0.42	0.54	0.44	0.34	0.52	0.33	0.30	0.45	0.47	0.43
$p(F_{current}> F_{LIMIT})$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$S_{current}/S_{MSY\_d}$	0.49	0.68	1.17	0.64	1.09	1.35	0.78	1.03	1.71	0.96	1.03	1.06
$p(S_{current}<S_{MSY\_d})$	1.00	1.00	0.23	1.00	0.31	0.17	0.99	0.43	0.06	0.59	0.44	0.39
$F_{current}/F_{MSY}$	1.68	1.36	0.94	1.38	0.95	0.78	1.27	1.04	0.65	1.07	1.01	1.00
$p(F_{current}> F_{MSY}$	1.00	0.98	0.38	0.99	0.37	0.15	0.95	0.60	0.02	0.64	0.53	0.49

Appendix 1.

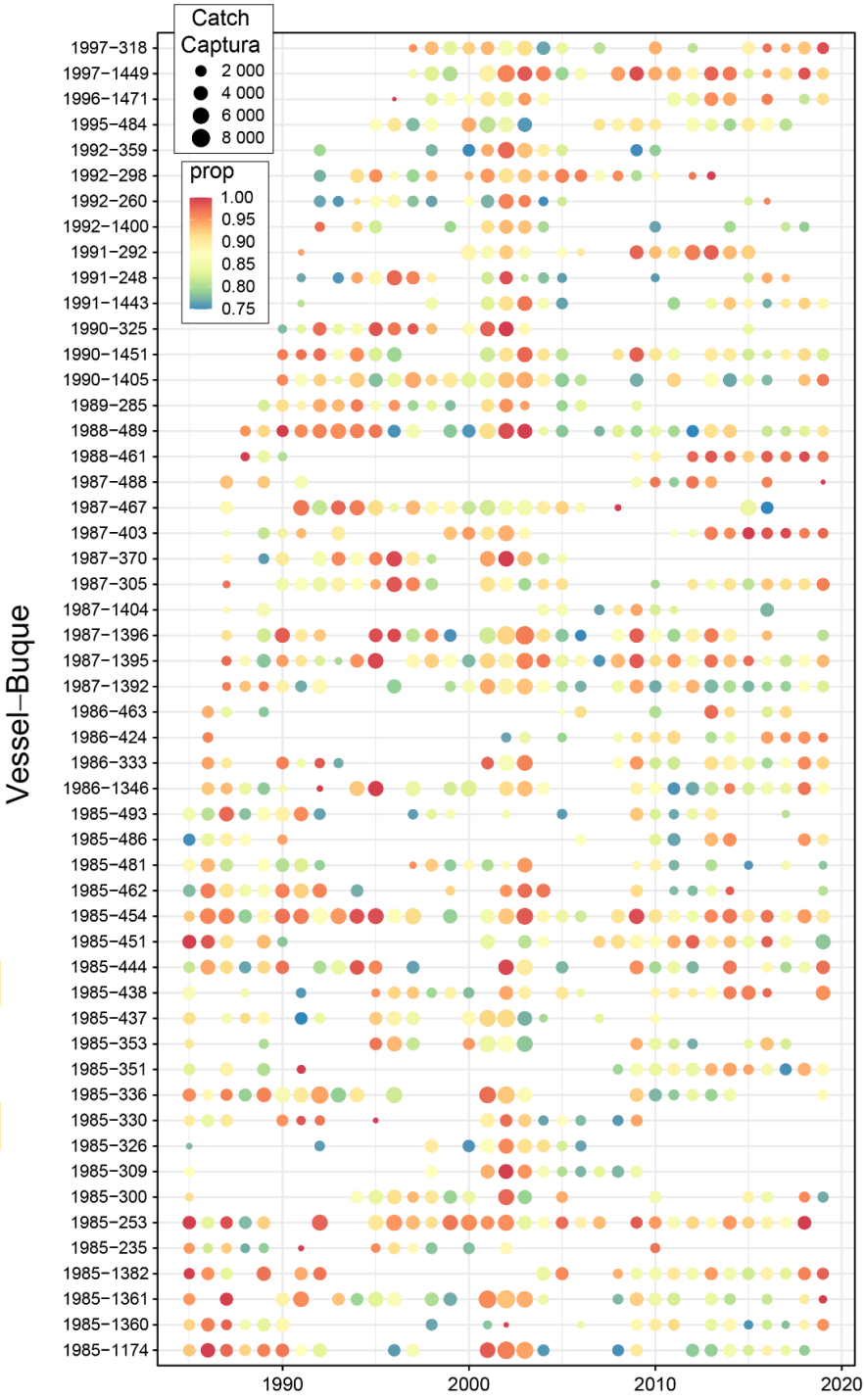

FIGURE A1. Vessels included in the standardization of CPUE for the index for the dolphin-associated fishery. Vessels are shown on the y-axis. The size of the dot represents the annual catch and the color the annual proportion of sets on dolphins.

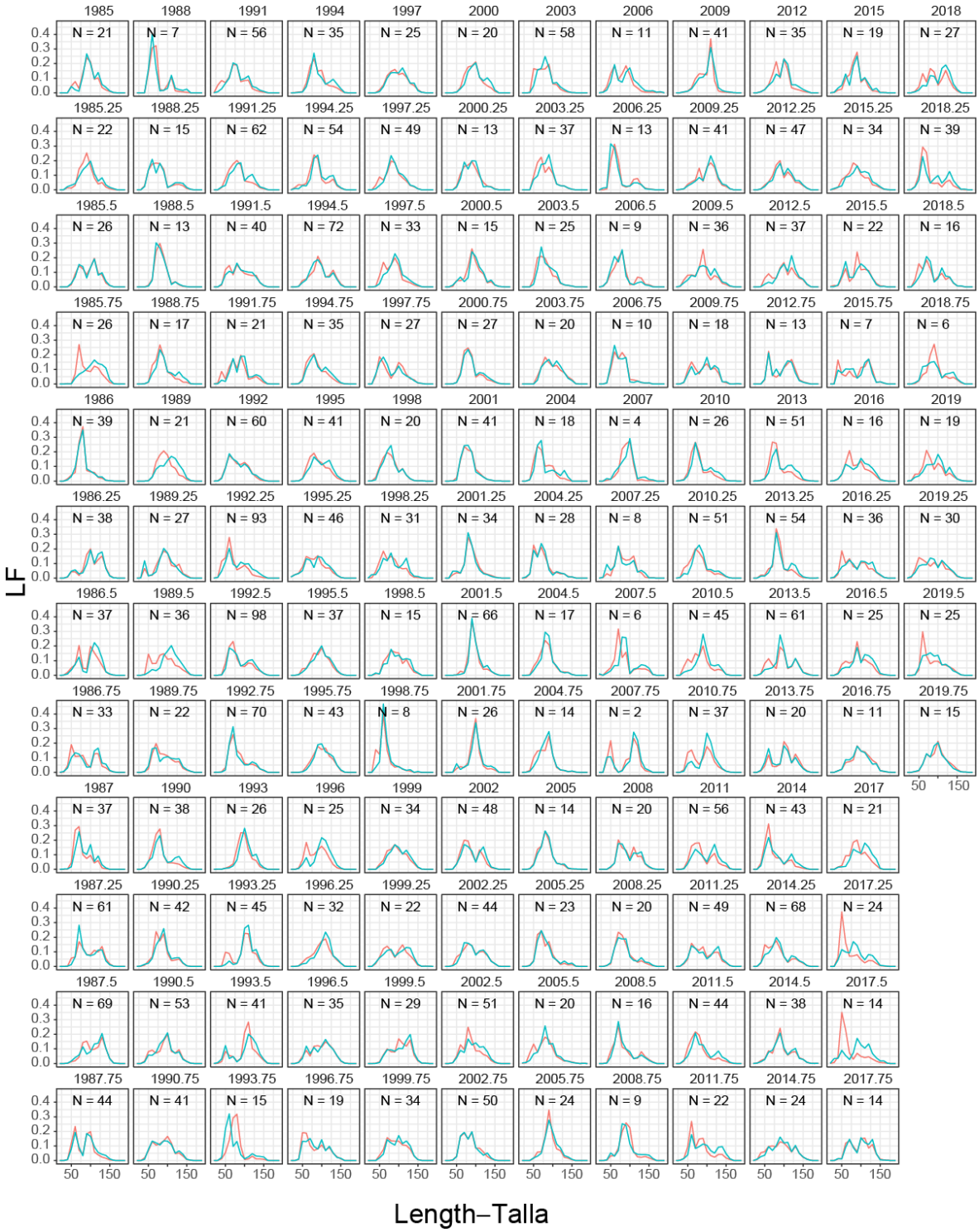

FIGURE A2. Comparison of nominal (red lines) and standardized (blue lines) length frequencies used to represent the index of abundance.

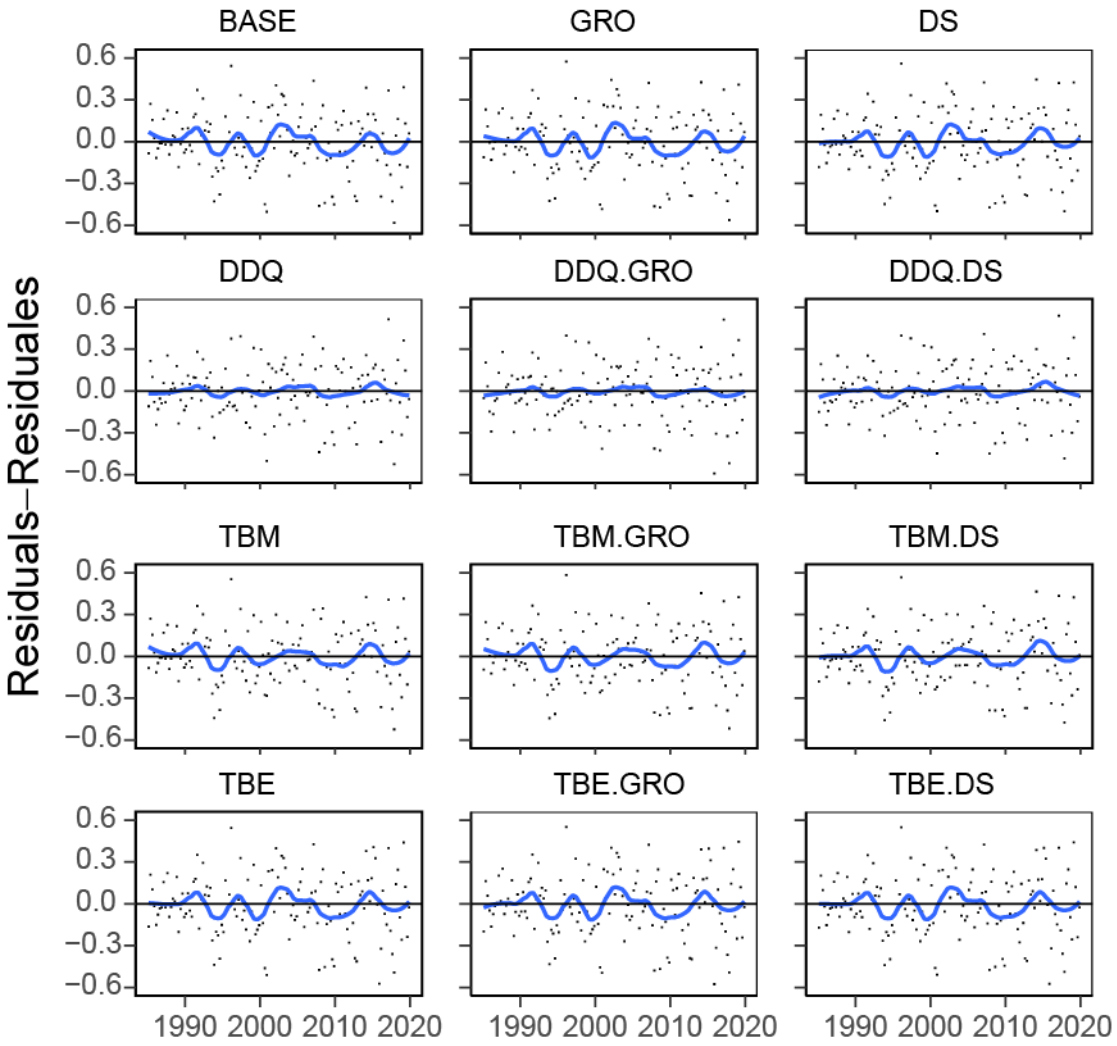

FIGURE A3. Residuals (log (observed index) – log (expected index)) for the twelve model configurations with steepness *h*=1. The lines were built using the *R stats::loess* function for fitting a local polynomial smoother with span=0.25.

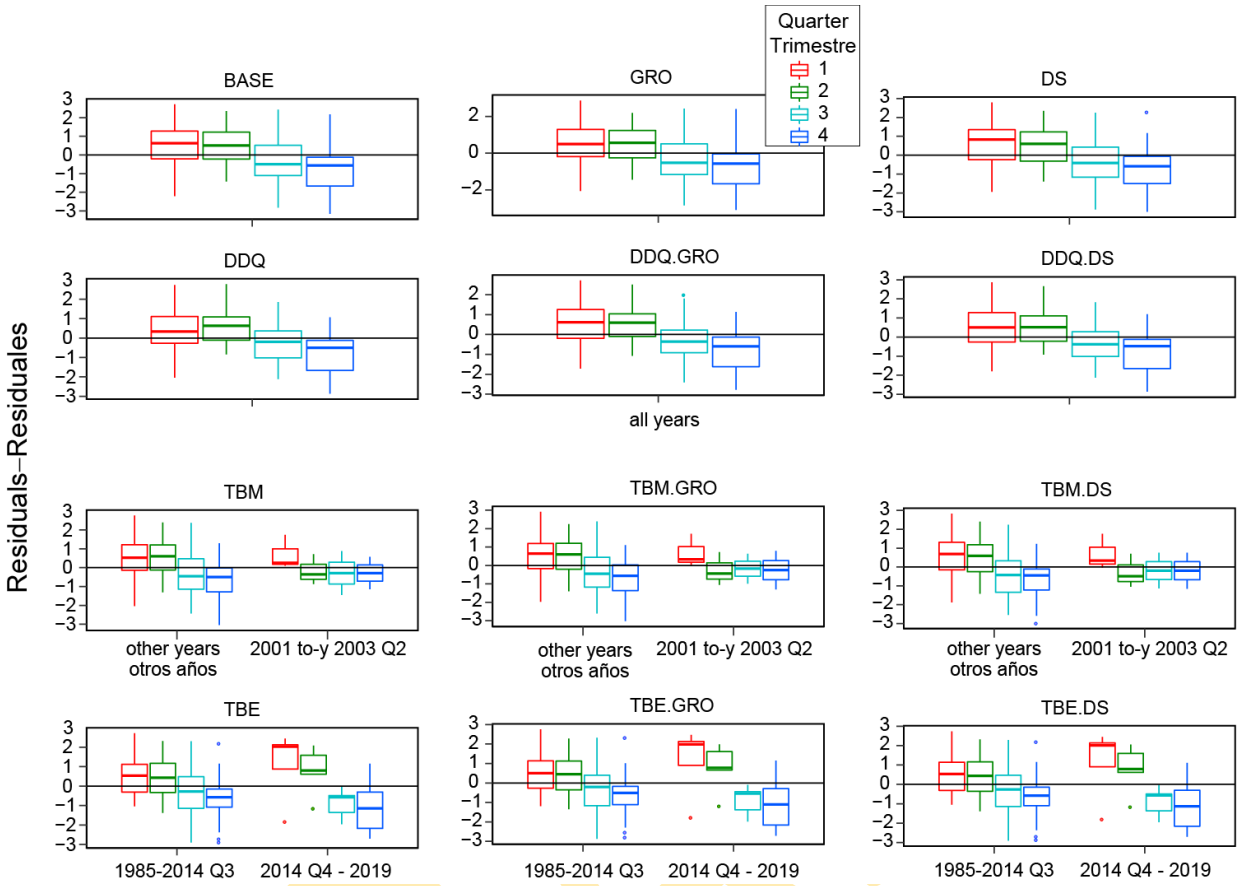

FIGURE A4. Residual (log (observed index)-log(expected index) plots for the twelve model configurations with steepness fixed at *h*=1.

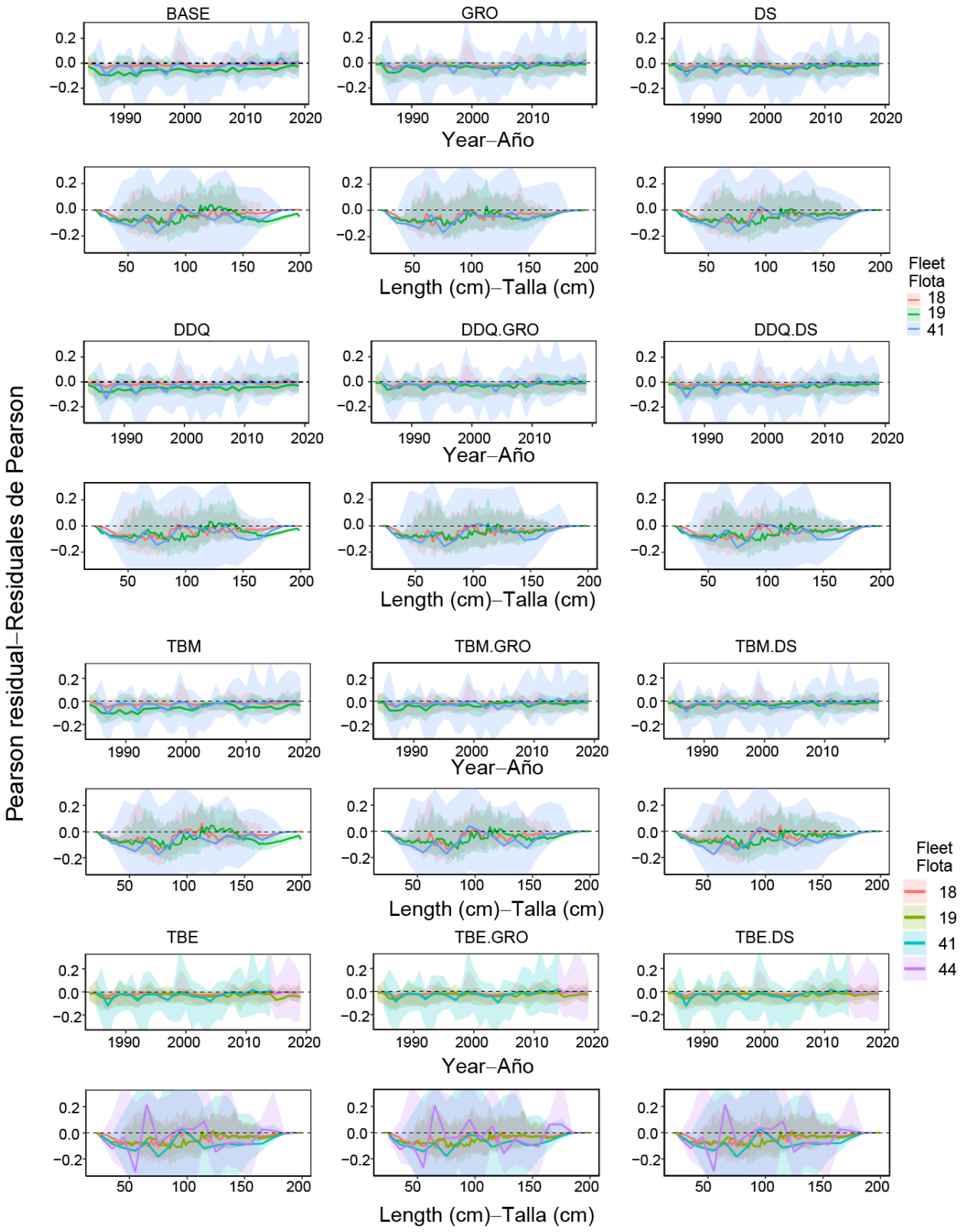

FIGURE A5. Residual plots for the survey (41 and 44) and two fisheries F18 and F19. The lines are the median residuals and the shaded area are the 25 and 75 percentiles.

Appendix 2.
INTEGRATED MODELS DIAGNOSTICS

Age-structured production model (ASPM): this diagnostic (Maunder and Piner, 2015) may be used to: (i) evaluate model misspecification, (ii) ascertain the influence of composition data on the estimates of absolute abundance and trends in abundance, and (iii) check whether catch alone can explain the trends in the indices of abundance. The ASPM diagnostic is computed as follows: (i) run the base case model; (ii) fix selectivity parameters at the maximum likelihood estimate (MLE) from the base case model, (iii) turn off the estimation of all parameters except the scaling parameters (R_0) , and set the recruitment deviates to zero; (iv) fit the model to the indices of abundance only; (v) compare the estimated trajectory to that of the base case. There is evidence of the existence of a production function if the ASPM can fit well the indices of abundance that have good contrast (*i.e.* those that have declining and/or increasing trends), it is also likely that the index, in combination with the catches, provides information on absolute abundance (Maunder and Piner 2015). When the catches cannot explain the changes in the indices, the ASPM will fit poorly. This can have several causes: (i) the stock is recruitment-driven; (ii) the stock has not yet declined to the point where catch is a major factor influencing abundance, (iii) the base-case model is incorrect, or (iv) the indices of relative abundance are not proportional to abundance. Checking whether the stock is recruitment-driven involves estimating recruitment deviations when fitting the model (ASPM-R). If this is still not able to capture the population trajectory estimated in the integrated model, it can be concluded that the information about scale in the integrated model is coming from the length composition data. Large confidence intervals on the abundance estimated by the ASPM also indicate that the index of abundance has little information on absolute abundance.

Catch-curve analysis (**CCA**) is done by fitting the integrated model only to the length composition data, and estimating all parameters except the auxiliary parameters associated with the index (Carvalho *et al*. 2017). The decline in the logarithm of the proportion of catch-at-age with age (the catch curve) provides information on fishing mortality (natural mortality assumed known), and when combined with catch data provides information on abundance. The CCA used to verify whether the temporal trend implied by the size composition data is consistent with that coming from the index of abundance. If the two trends are similar, then there is more confidence that the estimated abundance trend is accurate. Two variants of the CCA where used, one that is fit only to data from the fisheries and other that is fit only to the survey data.

Likelihood profile on the global scaling parameter: A likelihood profile of the average recruitment in an unfished (virgin) population in logarithm scale, lnR0, is used to determine whether information about absolute biomass scaling is consistent among data sets (*e.g.*, Francis, 2011;; Lee *et al*., 2014; Wang *et al*., 2014 2). The profile is done by fixing lnR0 to a range of values around the maximum likelihood estimate (MLE) and estimating all other parameters, then obtaining the contribution of each data set and penalty components to the likelihood conditioned of the value of lnR0. The profile quantifies how the fit to each data component is degraded by changing the population scale. Data components with a large amount of information on population scale will show loss of fit (smaller likelihood, or larger negative-log likelihood) as population scale is changed from its best estimate (Lee *et al* 2001). If different data components favor different values for ln(R0), there is contradictory information among them, conditioned on the model, thus pointing to potential model misspecification.

Retrospective analyses: these analyses are useful for determining how consistent a stock assessment method is from one year to the next (Mohn, 1999). The analysis is generally done by eliminating data for the last time step, then repeating the model fit without changing the method and assumptions, removing the last and the second last, running the model again and so on, until a desired amount of data is cumulatively removed. This shows the effect on the resulting estimated quantities of including more data.

Inconsistencies in the results of this progressive removal of data are a signal of inadequacies in the assessment models. The assessment model has a quarterly time step, but new data are updated annually (four quarters at once). Thus, the retrospective analysis was done by removing whole years of data at once.

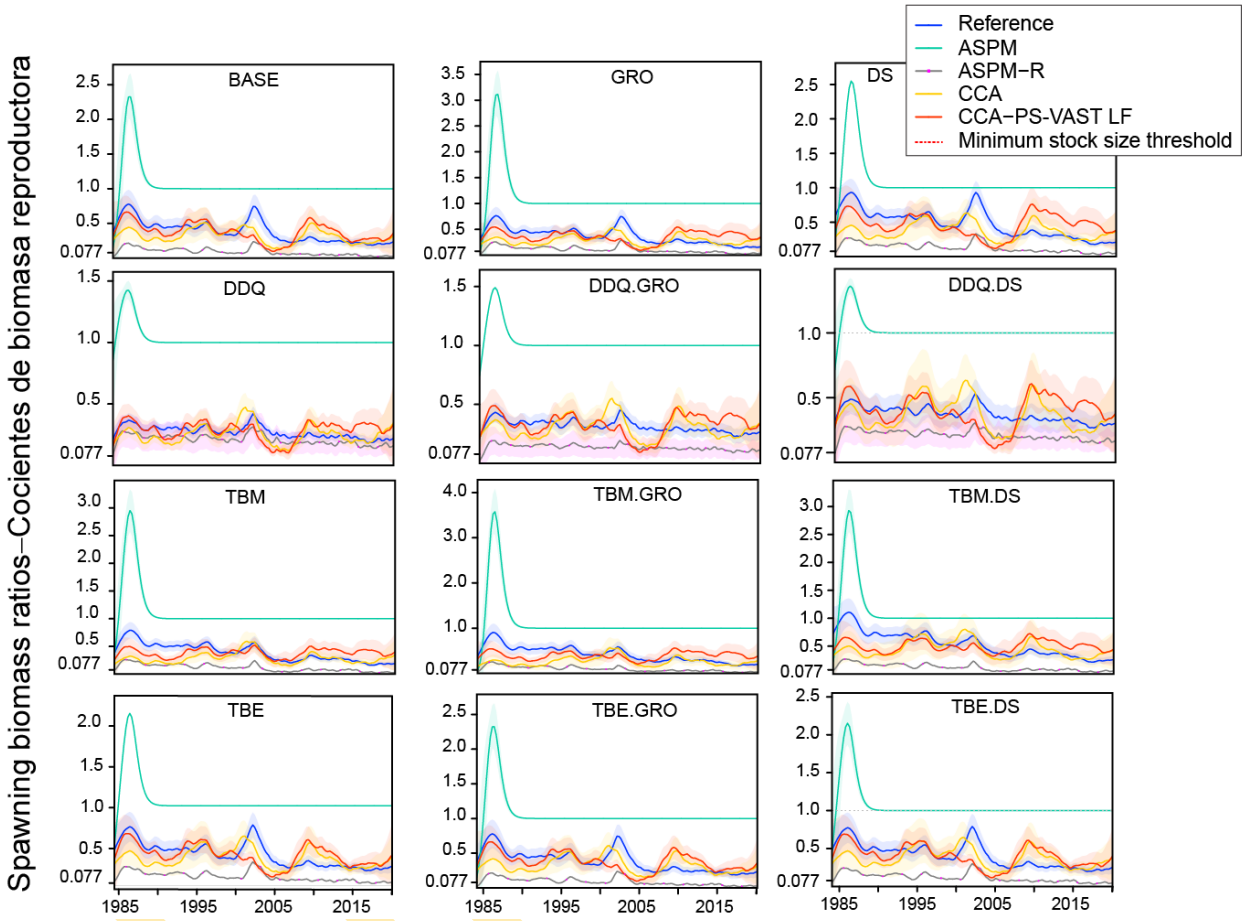

FIGURE A6. Comparison of estimated spawning biomass ratio of yellowfin tuna in the EPO between each reference model and the corresponding ASPM, ASPM-R, and CCA, and CCA-PSVAST models.

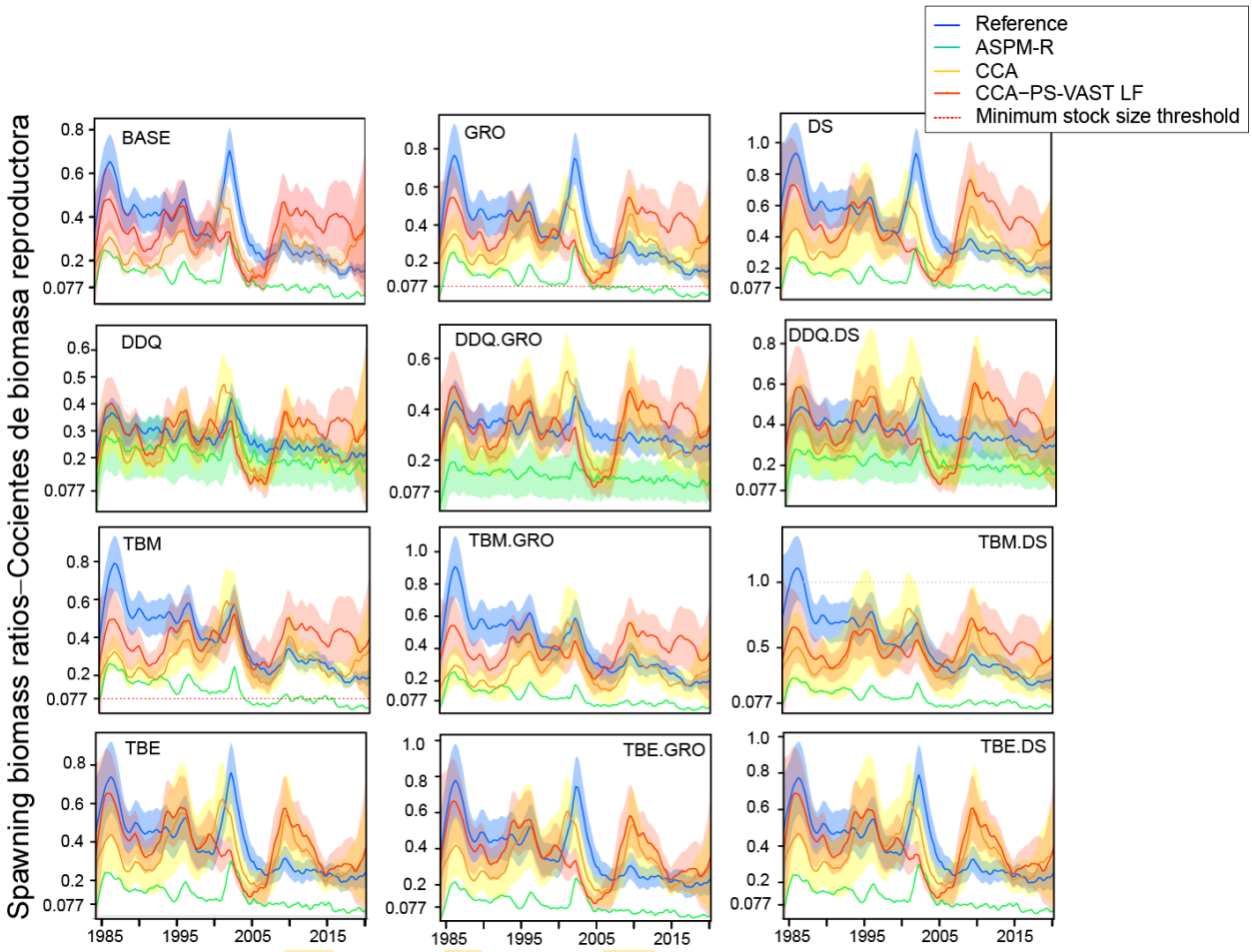

FIGURE A7. Comparison of estimated spawning biomass ratio of yellowfin tuna in the EPO between each reference model and the corresponding ASPM-R, and CCA, and CCA-PSVAST models. The lines are the MLE and the shaded areas the confidence intervals (CI). Models without CI did not produce a positive definite hessian.

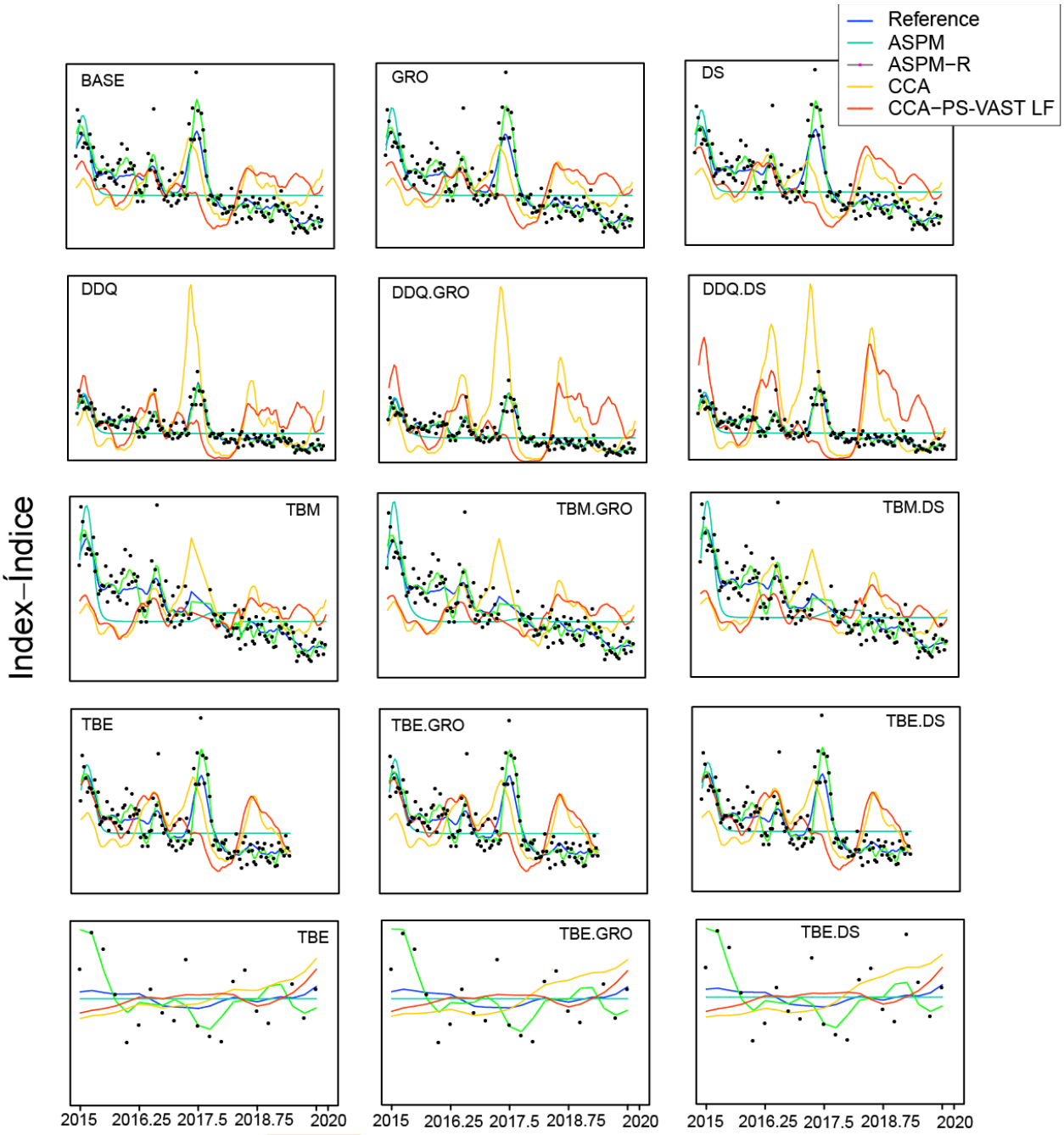

FIGURE A8. Comparison of fits and prediction of indices of abundance for yellowfin tuna in the EPO between each reference model and the corresponding ASPM, ASPM-R, and CCA, and CCA-PSVAST models. The lines are the MLE.

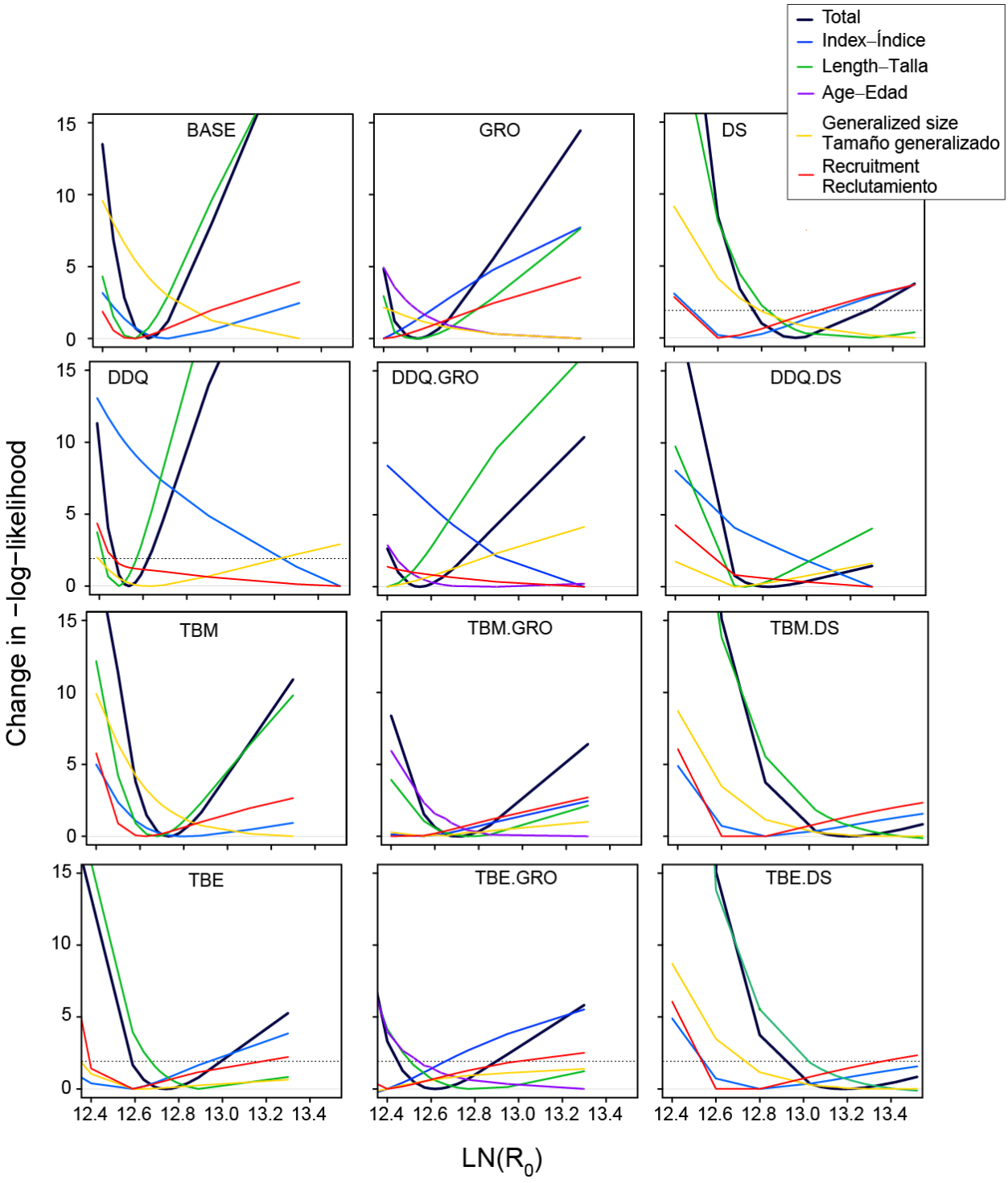

FIGURE A9. Comparison of the likelihood profile for lnR0 (scaling parameter) for the twelve reference models for yellowfin tuna in the EPO.

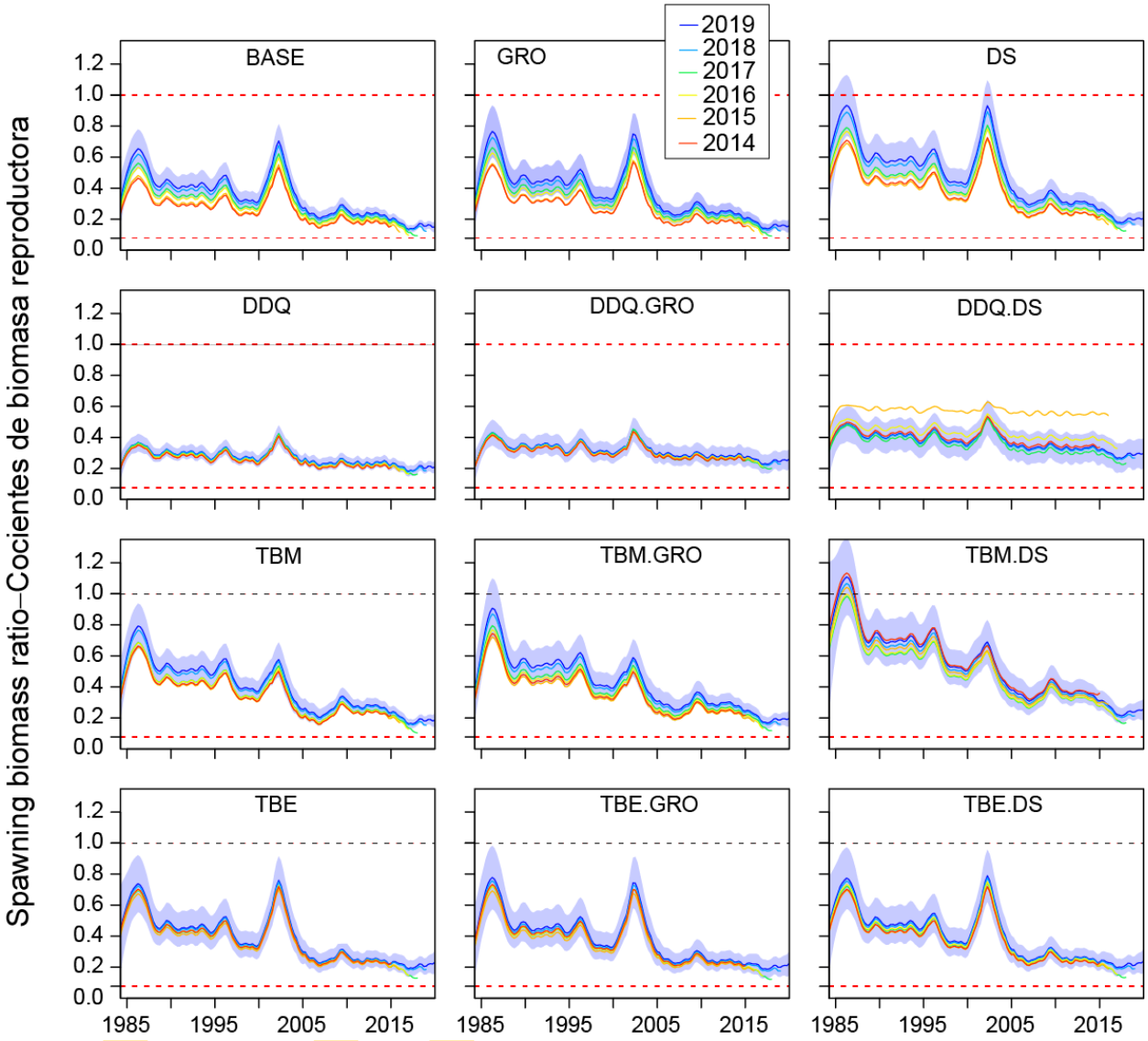

FIGURE A10. Retrospective patterns of spawning biomass ratio for the twelve reference models for yellowfin tuna in the EPO.

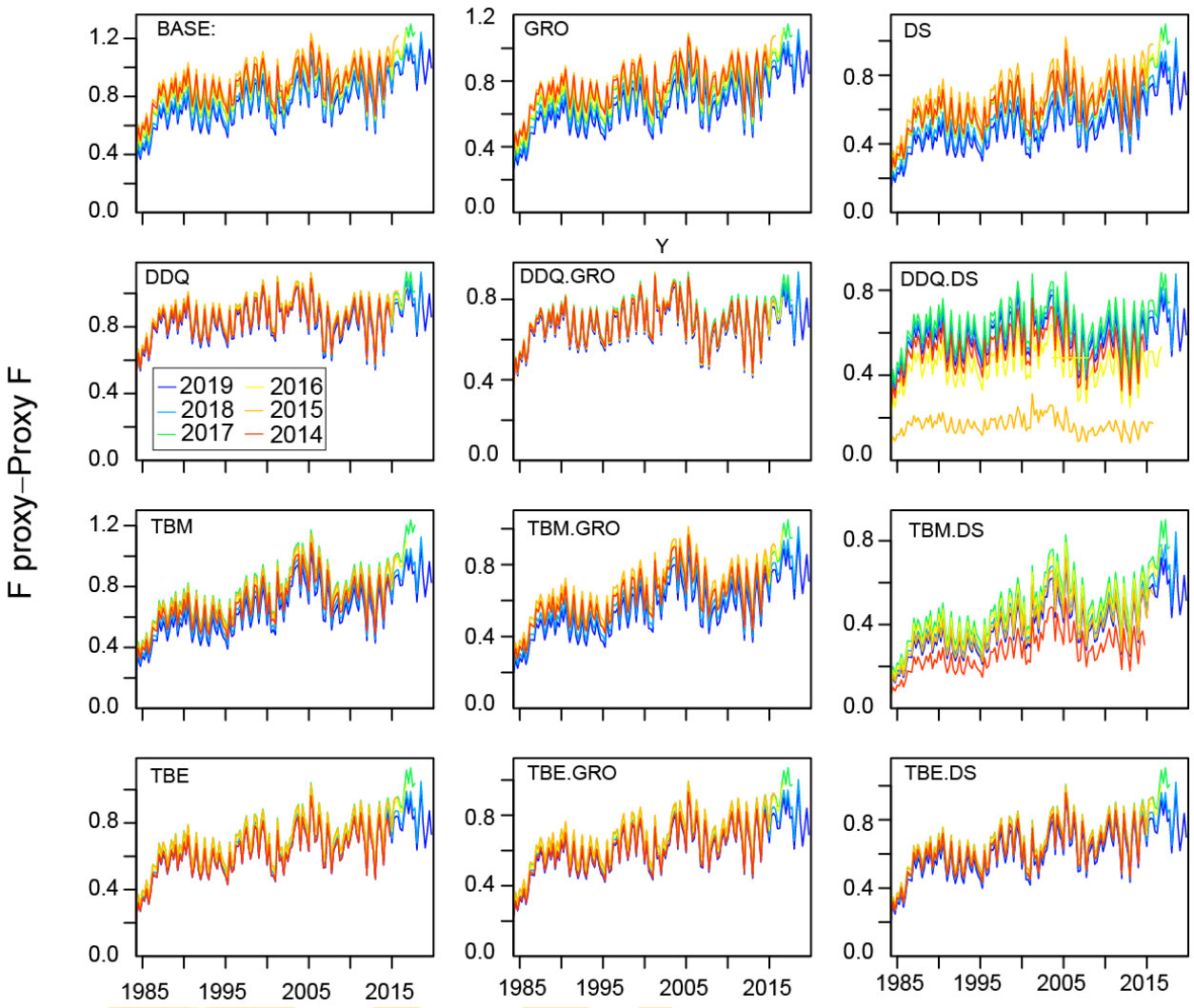

FIGURE A11. Retrospective patterns of fishing mortality proxy (1- SBR) / (1-SBR_{MSY}) for the twelve reference models for yellowfin tuna in the EPO.

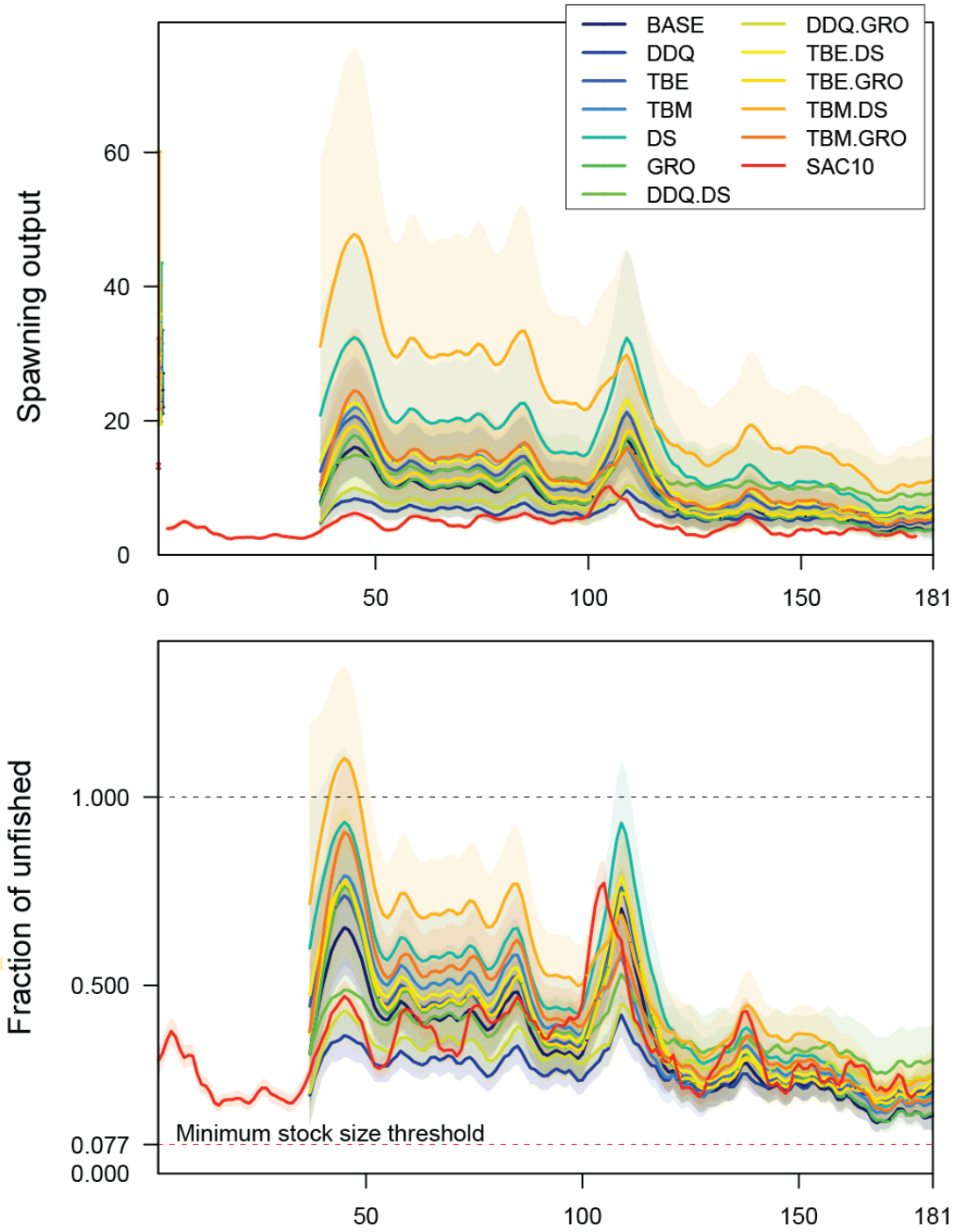

FIGURE A.12 Comparision with the SAC-10 assessment: (top) index of spawning biomass, (bottom) spawning biomass ratios (SBRs) for yellowfin tuna in the EPO for the 12 models from the reference set that have steepness 1 and for the SAC10 assessment. The solid curve illustrates the maximum likelihood estimates. The shaded areas indicates the approximate 95% confidence intervals around those estimates.The colored bars and points at the beginning of the panels area the estimates of virgin spawning biomass for each model. The red dashed horizontal line (at 0.077) identifies the SBR at the limit reference point.

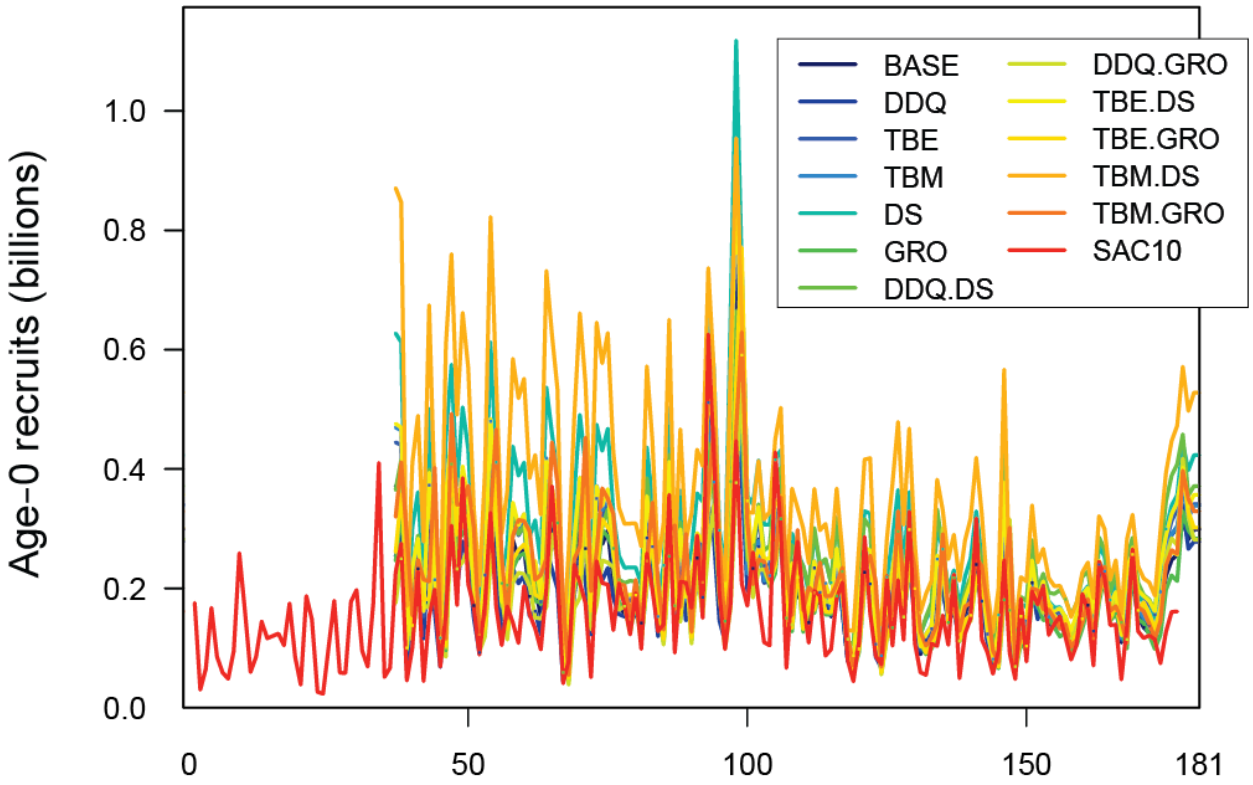

FIGURE A13. Estimated quarterly (A) recruitment of yellowfin tuna to the fisheries of the EPO from the 12 models with steepness *h*=1 and from the SAC-10 model. The bold line illustrates the maximum likelihood estimates (MLE) of recruitment.